THE UNIVERSITY
of EDINBURGH

Investigating the role of miR-21 in autophagy and VMP1

Kate O'Leary

Master of Science by Research

The University of Edinburgh

2016

Table of Contents

Table of Figures

Declaration

I hereby declare that this thesis has been composed by myself and that the work presented here is my own, unless otherwise stated, and has not been submitted for any other degree or professional qualification.

Kate O’Leary

Acknowledgements

My supervisors for this project have been Dr Elaine Nimmo and Professor Jack Satsangi. I would like to thank them for their support and supervision for the past four years. They have been wonderful supervisors, and their encouragement and guidance has been invaluable to me. I have been very lucky to have had them as my supervisors.

There are many in the lab I would also like to thank; Colette, for also being a kind and supportive presence in the lab, Nidhi, whose presence is sorely missed, Jacqueline for keeping me running, and Alex whose patience has been remarkable, an excellent R teacher!

Thank you to Stacey, for making my packed lunches and generally being an amazing flatmate! Thanks to Catriona for encouraging WhatsApp messages and asking me to be your bridesmaid! And James for the constant support, and the sunflowers.

And finally thank you to my parents, whose faith in me far surpasses my own.

Abbreviations

ATG16L1	Autophagy related 16-like 1
CARD	Caspase recruitment domain
CD	Crohn's disease
cDNA	Complementary DNA
Co-IP	Co-immunoprecipitation
dH20	Distilled water
DMEM	Dulbecco's modified eagles medium
DMSO	Dimethyl sulfoxide
EDTA	Ethylinediamenetetraacetic
FCS	Foetal calf serum
GAPDH	Glyceraldehyde-3-phosphate dehyrogenase
GI	Gastrointestinal
HC	Healthy control
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
LC3	Light chain 3
MDP	Muramyl dipeptide
miR-21	MircoRNA 21
NOD2	Nucleotide oligomerisation domain 2
PBD	Phosphate buffered solution
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RT qPCR	Real time quantitative PCR
SNP	Single nucleotide polymorphism
UC	Ulcerative colitis
VMP1	Vacuole membrane protein 1

Abstract

Previous work has identified a differentially methylated region (DMR) around the miR-21/VMP1 locus in IBD patients compared to HC and elevated levels of VMP1 and miR-21 have been detected in patients. MiR-21 is a microRNA that is associated with immune related disorders, and VMP1 Is a key regulator of autophagy, a pathway heavily implicated in CD pathogenesis.

The aim of this project was to elucidate the role of miR-21 in IBD pathogenesis and whether this is affected through VMP1. There were three main approaches to this question, firstly to establish the genetics of the region in CD, UC and HC and to correlate this to the expression of miR-21, VMP1 and methylation of the region. Secondly, using pyrosequencing to study methylation of this region, and 8 others to develop a diagnostic model for IBD. And thirdly to manipulate miR-21 levels in cell lines to investigate the effects on autophagy and VMP1.

Results: Investigating the genetics of the region so far has not found any mutations, possibly as this is a highly conserved region. Further sequencing is required before this can be fully established.

Pyrosequencing assays have been developed and performed on 8 CpG sites and a diagnostic model has been developed to distinguish IBD from HC, although it is not effective at distinguishing UC from and CD.

And finally, a VMP1 antibody for western blotting and miR-21 mimic and inhibitor for cell lines have been optimised. An interaction between VMP1 and NOD2 has been established by Co-IP.

Lay Summary

Inflammatory Bowel Disease (IBD) describes two diseases, Crohn's Disease (CD) and Ulcerative Colitis (UC). These diseases are caused by inflammation of the digestive tract and have debilitating symptoms. They are increasingly common worldwide, especially in westernised countries. Research into these diseases has shown the genetic differences between healthy people and patients, but now research has extended into the field of epigenetics. Epigenetics is the addition of different chemicals or proteins to genes that can change gene expression without changing the genetic code itself. This can be affected by the environment, and as we know the environment affects the chances of developing IBD it makes it an interesting area to study. An example of an epigenetic modification is methylation, when a methyl group is added to DNA.

This project studied a microRNA, miR-21. Previous work found that methylation was different at the miR-21 locus between people who were healthy and patients. This locus is also part of a larger gene, Vacuole Membrane Protein 1 (VMP1). VMP1 is involved in autophagy, a process where the cell is able to recycle its components, but also clear the cell of invasive particles such as bacteria. This process has been implicated in CD; we think that if autophagy isn't working properly the invasive particles persist in the gut cells for too long, making them more inflamed. MircoRNAs are small bits of RNA that help regulate gene expression and miR-21 has been implicated in inflammatory diseases.

To study this area more closely we sequenced this area in 300 people who were either healthy, had CD or had UC. This will help us discover any mutations that might be found in disease but not healthy people. We will also see if the levels of VMP1 and miR-21 in the cell are affected by these mutations. By studying the methylation at this point and other points in the genome, we may be able to create a test that can tell healthy people apart from patients which will help diagnosing people when they first get sick. And finally we will alter the miR-21 levels in cells in the lab and see if this changes autophagy in the cell or VMP1 levels, which will tell us if miR-21 is involved in this process.

1. Introduction

1.1 IBD

Inflammatory Bowel Disease (IBD) comprises two diseases, Crohn's disease (CD) and Ulcerative Colitis (UC). These are chronic diseases of the digestive tract that are seriously debilitating for sufferers. The two diseases are similar; symptoms for both include diarrhoea, bloating, weight loss, vomiting and severe abdominal cramps among them. They are caused by inflammation of the digestive tract in a dysregulated response to normal intestinal microbiota. CD can affect the digestive tract from mouth to anus although it primarily affects the ileum. The inflammation is transmural, and discontinuous. UC is characterised by inflammation in the colon which tends to be continuous and limited to the mucosal layer.

1.2 Incidence of IBD

Worldwide the incidence of IBD is increasing, and incidence and prevalence of the diseases are highest in the western world (Europe, North America, Australia and New Zealand). In Europe, for example, estimates of the incidence for UC range from 0.6-24.3 per 100,000 and estimates for CD range from 0.3 to 12.7 per 100,000 (Molodecky et al. 2012). As well as the personal burden there is an increasing financial burden on healthcare providers. In 2012 the annual cost of IBD in Europe was estimated to be as high as €4.6-5.6bn (Burisch et al. 2014). Although the incidence is highest in westernised countries the rate of incidence is rapidly increasingly in Asia (Ananthakrishnan et al. 2013).The industrialisation of countries has an effect on the incidence of IBD, due to a change in environmental factors and an increase in access to healthcare. Incidence therefore, will only continue to increase, causing IBD to be a global disease (Ng 2015).

1.3 Environment

Environmental factors are known to have a significant effect on the development of IBD. As has been mentioned, the industrialisation of countries increases the incidence of IBD. Several studies have shown that individuals who were living in countries with low incidence who move to countries of high incidence adopt the

same risk as those in their new country, demonstrating the significance of the environment on the development of these diseases (Benchimol et al. 2015; Probert et al. 1993).

1.3.1 Smoking

Smoking is the most significant contributing environmental factor in IBD; interestingly, although smoking is a risk factor for CD, it is protective against developing UC. Many studies have shown an increased risk of CD in current and ex-smokers with an odds ratio of up to 2.0 (Calkins 1989). Smoking also has an effect on disease course; patients with CD who smoke are at a 2.5 times higher risk of surgical recurrence (Reese et al. 2008) and relapse more often (Nunes et al. 2016). People who smoke however, are less likely to develop UC (Harries et al. 1982), and for patients with UC smoking can prevent recurrence of flare ups (Beaugerie et al. 2001) and decrease the chances of a colectomy (Szamosi et al. 2010). However, upon cessation of smoking the risk of developing UC increases to that greater than that for a non-smoker (Higuchi et al. 2012). While it isn't clear exactly how smoking affects the development of IBD smoking is known to affect cytokine levels, T cell proportions, the mucus membrane of the gut (Allais et al. 2015), the gut microbiome (Biedermann et al. 2013) and can cause epigenetic changes in the blood (Zeilinger et al. 2013).

1.3.2 Appendectomies

Appendectomies have also been shown to explain IBD. They seem to be protective against UC (Russel et al. 1997), and there is some conflicting evidence as to whether they are a risk factor for CD. It is thought that microbiome dysbiosis in the appendix may be a priming event for UC and therefore appendectomies may reduce risk of UC (Roblin et al. 2012). Additionally, an altered CD4/CD8 ratio has been found in the appendix, also suggesting it may be a priming site for UC (Matsushita et al. 2005).

1.3.3 Hygiene hypothesis

The 'hygiene hypothesis' has also been postulated to explain IBD. The development of the immune system is an important stage in a child's life and it is affected by the environment (Hoffjan et al. 2005). As countries develop and there is better access to clean water and increased sanitation infectious diseases rapidly decline, improving public health but also decreasing the likelihood of contact with infectious agents. Autoimmune diseases such as diabetes and IBD are a disproportionate immune response to normal conditions, to commensal flora in the case of IBD. As the immune system is trained by exposure it is thought that a lack of exposure or a lack of diversity in bacteria inadequately trains the immune system (Koloski et al. 2008).

1.3.4 Diet

Finally, diet is known to be a significant contributing factor. The western diet which is high in fat, low in fibre is a risk for IBD (Burisch et al. 2014). A study in women showed that long term consumption of soluble fibre, particularly fruit, could reduce the chances of CD by up to 40% (Ananthakrishnan et al. 2013). Children with IBD consumed a higher proportion of sugar and soft drinks (Jakobsen et al. 2013). Diet directly affects the microbiome composition and function, in a way that can promote inflammation (Hwang & Weiss 2014). Diet factors therefore can cause long term changes to the microbiome, contributing to the likelihood of disease.

1.4 Microbiome

A dysregulated microbiome is found in patients with IBD. The human gut is home to more than 1000 microbial species, and each individual is expected to have at least 160 different types (Balzola et al. 2010a). These microbes, in their millions, have an important symbiotic relationship with their host, providing important metabolic functions. However, it is also very important that the microbiome remains non-pathological and does not breach the intestinal wall. A dysregulated microbiome has been associated with many diseases, such as IBD, but also asthma, obesity and cardiovascular disease, and it also has a significant role in metabolism, behaviour, the immune system and even the development and maturation of the digestive system itself. The microbiome and the innate immune system interactions are known to be critical for disease (Thaiss et al. 2016).

1.5 Treatment

Current treatment is limited to immunosuppressants, such as azathioprine, biologics, such as infliximab, an anti-TNFa antibody, steroids and surgery. These can help to alleviate symptoms but are not a cure, and have serious side effects. The aetiology of IBD is not fully understood although research has shown there is a significant genetic component to the disease. It is thought that environmental factors affect genetically susceptible individuals causing them to react excessively to commensal bacteria in the gut thus prolonging inflammation. Establishing the full aetiology of the disease is important for developing better treatments for these serious and debilitating diseases.

1.6 Genetic studies

Twin and familial studies showed that there was a strong genetic component to these diseases. The strongest risk factor for disease is having a family member with the disease (Orholm et al. 1991). Monozygotic twins also have a higher concordance than dizygotic, demonstrating a genetic contribution to disease (Halme et al. 2006). Early studies used linkage analysis which identified two IBD risk loci, IBD1 and IBD9 (Lander & Kruglyak 1995), IBD1 was shown to be caused by the susceptibility gene NOD2 (Ogura et al. 2001). Linkage studies however, are not capable of detecting multiple genes of small effect and so genome wide association studies (GWAS) were later used to investigate the genetics of IBD.

By using thousands of cases and controls they are able to identify single nucleotide polymorphisms (SNPs) that are associated with disease. These SNPs may not be causative themselves but tag a region of DNA that is involved in disease pathogenesis. These studies identified 200 SNPs, of which 148 are associated with susceptibility to IBD, and the remainder are split between UC and CD (J. Z. Liu et al. 2015). This contributes to 13.1% and 8.2% of susceptibility to CD and UC respectively.

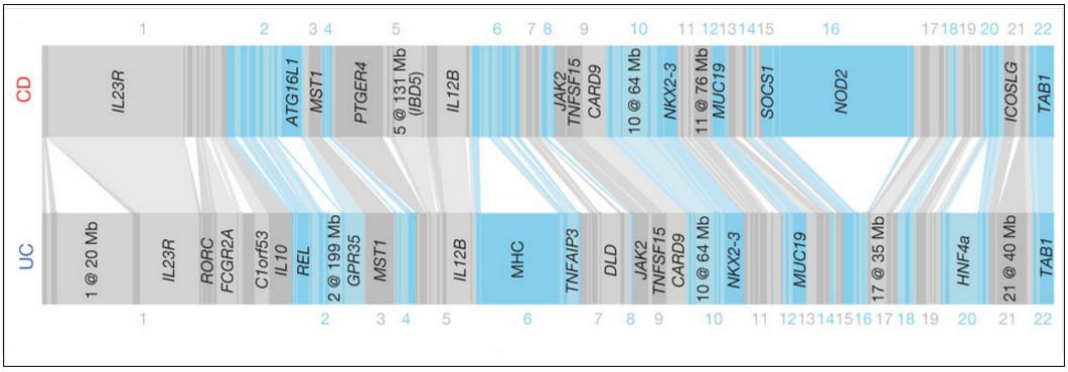

Figure 1: Schematic adapted from Jostins et al, 2012 showing variance of disease explained by that locus. Lines between the two diseases indicate a shared susceptibility locus

There is a still a significant proportion of heritability unaccounted for. This may be in part because of weaknesses in the GWAS approach- it can't identify larger polymorphisms such as indels, it doesn't account for epistatic interactions and the focus on common variants –, or an overestimation of heritability from twin studies in the first place, but it has also been proposed that this missing heritability could be caused by epigenetics.

1.7 Epigenetics

Epigenetics describes factors that alter gene expression without altering the genetic code itself. These may be by physical alterations the genome, such as the addition of a methyl group to a CpG site, or acetylation of the histone molecules that with DNA form chromatin. It also describes the effects of small RNA interactions such as sRNA and microRNAs.

1.7.1 MicroRNAs

MicroRNAs are small (~20 nucleotides) non-coding RNA fragments that bind to mRNAs and mark them for degradation. They are an important regular of gene expression; by binding to mRNAs they can physically prevent translation of the protein, and can also mark the mRNA for degradation. They are transcribed from introns and sometimes exons, often independently from their host genes. This is the primary transcript, which can be several thousands of kilobases long. This is

processed by the RNAse III enzyme, DROSHA to form the pre-miRNA which is then exported from the nucleus. The pre-miRNA forms a hairpin loop of about 75nt long. In the cytoplasm it is further processed by the RNAse III enzyme DICER to form the mature miRNA, between 18-24 nt long (Winter et al. 2009). The mature miRNA binds to argonaute proteins and guides them to mRNAs where they have complementary sequences to the mature miRNA causing the mRNA to be degraded. In this way microRNAs are regulators of gene expression.

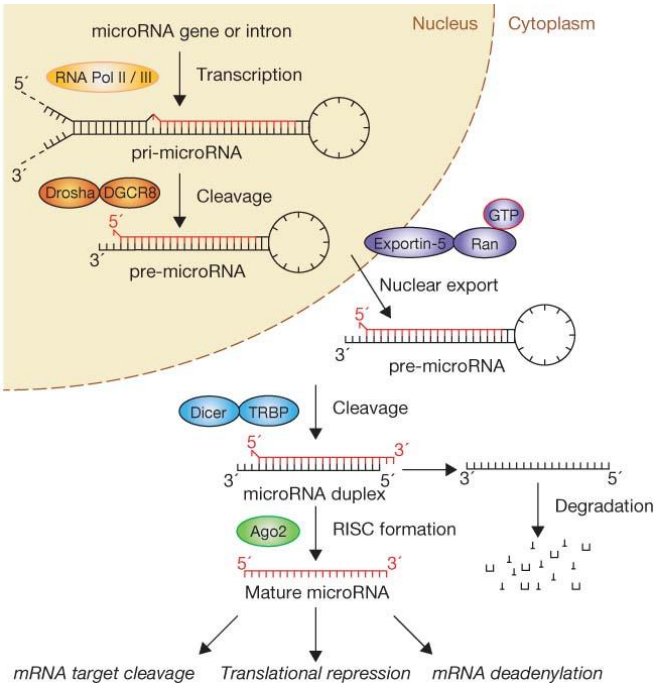

Figure 2: Schematic adapted from Winter et al 2009 demonstrating the processing of microRNAs

1.7.2 Methylation studies

Mammalian DNA can be methylated at the cytosine in a CpG sites- where a cytosine is followed by a guanine. 5'-methyltcytosine is prone to deamination, causing a mutation to thymine. As such, CpGs occur less often than would be expected in the genome. However, there are areas of DNA that have a higher proportion of CpGs and these tend to be unmethylated. These are denoted as CpG islands (CGI) and are found at gene promoters (Deaton & Bird 2011). DNA methylation is associated with decreased gene expression (Bird 2002), which is caused by physically impeding

transcription factors from binding, but also by recruiting chromatin remodelling proteins.

Methylation chips are a useful way of observing methylation at numerous CpG sites across the whole genome. First the DNA is treated with sodium bisulphite (hereafter referred to as bisulphite conversion) which causes unmethylated cytosines to be converted to uracils. The DNA is then fragmented and hybridised on the chip- each CpG site will hybridise to one of two 50nt DNA probes- one for an unmethylated CpG, or the other if it is methylated.

An epigenome wide methylation study was undertaken by the lab in order to investigate methylation differences associated with CD in children (Adams et al. 2014). It used an Illumina 450K platform to measure methylation levels in leukocytes in newly diagnosed paediatric patients with CD and a replication cohort of paediatric cases with established CD. It identified 65 CpG sites that were associated with disease, including a differentially methylated region at the VMP1/miR-21 locus.

1.8 Autophagy

Autophagy describes the process whereby the cell clears itself of unnecessary subcellular components in times of stress. It is also used to help clear the cell of invasive particles, known as xenophagy (Bauckman et al. 2015). Autophagy is an important process that can help maintain homeostasis in the cell, even in times of nutrient starvation (Russell et al. 2014). This process starts with the formation of an autophagosome, an organelle that engulfs the target of autophagy (Tooze & Yoshimori 2010). The autophagosome then fuses with a lysosome, forming an autolysosome, which can then digest the contents of the autophagosome (Eskelinen & Saftig 2009; Kaur & Debnath 2015).

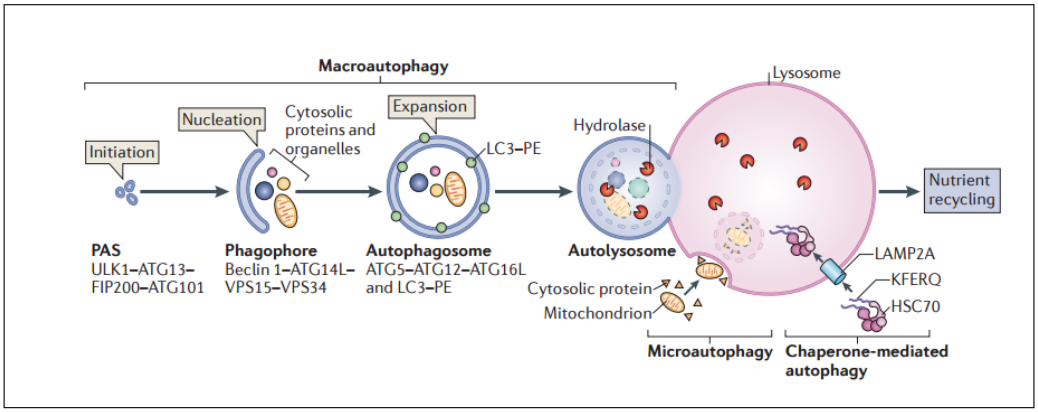

Figure 3: Schematic adapted from Kaur and Debnath, 2015, demonstrating autophagy

Autophagy has been implicated in IBD through genetic studies, for example a variant of ATG16L1 gene, (T300A) is strongly associated with CD (Murthy et al. 2014). This variant makes the ATG16L1 protein significantly more susceptible to caspase 3 degradation. As ATG16L1 is crucial for the formation of the autophagosome this diminishes basal autophagy and xenophagy, thereby increasing inflammation in the cell. A GWA study also identified SNPs near the IRGM gene that are strongly associated with CD and upon further investigation were discovered to be linked to a 20kb deletion upstream of IRGM. IRGM is important for the assembly of complexes that initiate autophagy (McCarroll et al. 2008). Finally, NOD2, the strongest genetic determinant of CD is known to interact with ATG16L1 and this interaction is crucial for the autophagic responsive to bacteria in the cell (Balzola et al. 2010b).

If autophagy is defective or delayed the cell is unable to respond to invasive pathogens, allowing them to persist in the cell, thereby driving up inflammation.

1.9 VMP1

VMP1 is a crucial regulator of autophagy; autophagy is blocked in the absence of VMP1, and its expression can trigger autophagy even in the presence of plentiful nutrients (Ropolo et al. 2007). VMP1 is expressed upon stimulation of autophagy triggers, such as rapamycin and starvation (Kang et al. 2011). VMP1, which is a transmembrane protein, is part of the Beclin 1-Class III PI3K complex, which is required for its recruitment to the autophagosomal membrane, initiating autophagy

(Molejon et al. 2013). The interaction of the VMP1-AtgD domain and the BH3-Beclin-1 domain allows the localisation of the complex to the pre-autophagosomal structure, allowing further autophagy specific proteins to be recruited to the autophagosomal machinery.

1.10 NOD2

Mutations in NOD2 are the strongest genetic determinant of CD (it is not associated with UC) (Cuthbert et al. 2002). NOD2 is a intracellular pattern recognition receptor- its key role is to recognise MDP, a component of bacterial cell walls, triggering an inflammatory, anti-microbial pathway (Grimes et al. 2012). It is also known to be involved in autophagy; stimulation of NOD2 activates autophagy in a pathway involving ATG16L1, another CD susceptibility gene (Cooney et al. 2010).

1.11 MiR-21

MiR-21 was one of the first microRNAs to be discovered. Its targets include many tumour suppressors so its upregulation has been associated with many kinds of cancer, including colorectal cancer (Asangani et al. 2008). The primary transcript for miR-21 begins in the 3' end of the VMP1 gene, and is transcribed in the same direction (Kumarswamy et al. 2011). It is independently transcribed (Fujita et al. 2008) but there is also evidence that a mature miR-21 is processed from an extra-long VMP1 transcript (Ribas et al. 2012).

1.10.1 MiR-21 and Disease

MiR-21 is involved in many diseases, most notably cancers. However, it is also dysregulated in inflammatory diseases. It is upregulated in patients with diabetes (Osipova et al. 2014), psoriasis (Meisgen et al. 2012) and MS (Fenoglio et al. 2011) indicating a significant role in maintaining inflammation. It is upregulated in both gastric and colonic cancers. It can be detected in serum, urine and stool and is already being as a biomarker in these (Toiyama et al. 2013).

1.10.2 MiR-21 and Autophagy

There is some evidence to say that miR-21 is involved in autophagy. A study investigating renal ischaemia-reperfusion injury found that Rab11a was a target of miR-21. Rab11a is involved in autophagy-it facilitates the fusion of endosomes and autophagosomes- and so autophagy is inhibited on overexpression of miR-21 (X. Liu et al. 2015). Cells treated with antimiR-21 were also reported to have an increase in the autophagy genes LC3-II and beclin-1. Finally, downregulation of miR-21 causes the inhibition of mTORC1 activity- mTORC1 is a regulator of autophagy (Sha et al. 2015). There is no evidence yet as to whether miR-21 has an effect on its host gene, VMP1.

1.10.3 Mir-21 in Inflammation

MiR-21 has a complicated role in inflammation. It is thought to be involved in the switch from a pro-inflammatory response to an anti-inflammatory response. MiR-21 is a marker of immune cell activation, for example in T cell differentiation (Wu et al. 2007) , but its delayed activation suggests it may also be involved in the resolution of inflammation (Sheedy 2015).

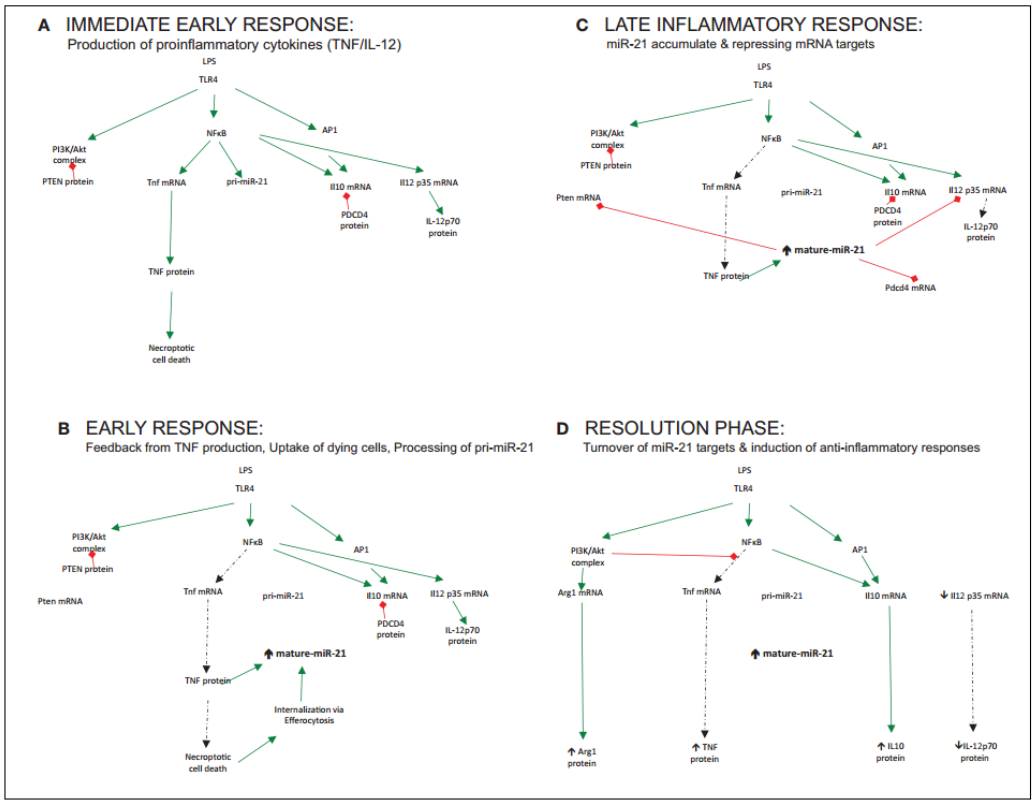

Figure 4: Schematic adapted from Sheedy 2015 demonstrating the complicated role of miR-21 in inflammation. 'A) Immediate early response: production of proinflammatory cytokines (TNF/IL-12). (B) Early response: feedback from TNF production, uptake of dying cells, processing of pri-miR-21. (C) Late inflammatory response: miR-21 accumulates and repressing mRNA targets. (D) Resolution phase: turnover of miR-21 targets and induction of anti-inflammatory responses.'

1.12 Conclusion

The full aetiology of IBD is still obscure, although there are strong genetic determinants. This does not explain all of the estimated heritability, however, and much research has focussed on trying to explain this. We know that the environment has a strong influence of likelihood of developing disease and there are epigenetic changes in IBD. Epigenetics is influenced by the environment and it is hypothesised that the interplay between environment and epigenetics in genetically susceptible individuals can explain the development of disease.

Epigenetic markers can serve as biomarkers for disease and can be useful for earlier diagnosis, or tailored treatment plans for individuals with disease. They may also help to elucidate pathways and mechanisms involved in the disease, leading hope to improved treatments.

A previous study by our laboratory has led to the mir21/VMP1 locus for further investigation. It is a DMR and there are genetic links to the area too (which SNPS, which study). VMP1 is a key regulator of autophagy, a pathway already heavily implicated in IBD pathogenesis. Autophagy is a pathway by which the cell can clear itself of invasive particles- it is thought that if the cell is unable to do so effectively there will be an increase in inflammation. Mir21 has also been implicated in autophagy- it seems an increase in mir21 causes an inhibition of autophagy.

The role of mir21 in autophagy is to be further explored, and whether it mediates its effects through its host gene, VMP1.

2. Materials and Methods

2.1 Cell Culture

2.1.1 Maintenance of cells

Cells were cultured at 37 °C and 5% CO2 in a Galaxy 170S incubator. THP1 cells were cultured in RMPI 1640 (ThermoFisher) media and HEK293 cells with DMEM (ThermoFisher) supplemented with 100 units/ml Penicillin and 100ug/ml Streptomycin and FCS at a concentration of 10%. All experiments and passaging of cells was performed in a sterile hood with laminar air flow. All equipment used was sterile and cells tested for mycoplasma once a month. Cells were counted prior to experiments using a Cellometer X1 (Nexcelom).

2.1.2 Passing Cells

Cells were passed when they had reached ~80% confluency. Adherent cells were washed with PBS after media had been aspirated. ~2 mls of Trypsin-EDTA (0.05%), phenol /red (ThermoFisher) was added to the flask and cells incubated at 37°C for 5 minutes. Media was added to quench the trypsin and cells diluted into more media at 1/10 ratio. Suspension cells were also passed at a similar ratio, by aspirating 1ml of media and diluting into 10mls media for a 25cm flask.

2.1.3 Transient Transfection

Cells were cultured in a 24 well plate and seeded at a density of 5 X 10^{-5}. THP1 cells were electroporated using the Nucleofector (Lonza) and using the Nucleofector Kit V. HEK293 T cells were transfected using lipofectamine 2000 (life Technologies). A negative control for the mimic was used (AllStars Negative Control siRNA, Qiagen) at 5nMr for 48hrs. A negative control for the inhibitor was used (a random sequence with no homology to any mammalian genes, Qiagen) was used at a concentration of 50nM for 48hrs.

2.1.4 Autophagy assay

Cells were stimulated with bafilomycin (Invitrogen) at a concentration of 20nM for three hours. They were either fed with normal media or starved using a media with

no amino acid supplement. Cells were imaged using a confocal microscope and western blotted for LC3.

2.2 Protein analysis

2.2.1 Sample Preparation

Cell pellets were collected from 24 well plates by trypsining adherent cells or removing media with suspension cells. Pellets were lysed with 100-300ul (depending on pellet size) with a homemade lysis buffer. Total protein amount was quantified using the DC Protein Assay (Bio-Rad). 2X Lammelli bufer (BioRad) was added (supplemented with B-mercaptoethanol, 50ul +950ul Lammeli buffer) and sonicated at 10micron amplitude for 30s. The sample was then heat shocked for 5 minutes at 95°C.

2.2.2 Western Blotting

30ug of total protein was loaded onto a 10% or 15% SDS-PAGE gel with a protein ladder (Page Ruler, ThermoFisher). The tank was filled with running buffer the gel run at 80V for 15 minutes and then 120V for a further 40 minutes, until the ladder was clearly separated on the gel.

Protein was transferred to a PVDF membrane using the Invitrogen system. The membrane was washed first in methanol. Sponges and filter paper were soaked in transfer buffer before assembly. The protein was transferred for 90 minutes at 400mA on ice and the tank was filled with the transfer buffer.

2.2.3 Immunoblotting

Membranes were blocked after transfer in 5% BSA solution and then probed in 5% milk solution at an antibody concentration of 1:1000 overnight at 4°C. The membrane was washed 3 times with PBS-tween for 5 minutes and then probed with HRP conjugated secondary antibody for 1 hour at room temperature. The membrane was washed 3 times again with PBS-tween and treated with an ECL kit (GBI labs) and then visualised using the Li-COR Odyssey machine.

Probe	Antibody
VMP1	TMEM49/VMP1 (D1Y3E) Rabbit #12929, Cell Signalling Technologies
LC3	Anti-LC3 ab produced in rabbit, L7543, Sigma-Aldrich
Actin	B-Actin, (AC-15), sc-69879, Santa Cruz
HA	Anti-HA (Y11), Santa Cruz
Secondary antibodies	HRP-linked secondary, Dako

2.2.4 Co-Immunoprecipitation

30ul of protein G agarose beads were used per IP. Beads washed in lysis buffer (suspended in 1 volume lysis buffer, centrifuge for 13,000 rpm for 15 mins). 50ul of cell lysate was removed and stored for input. The beads were added to the remaining lysate and 1ul of antibody per IP. Left rotating overnight at 4 °C. The nest day beads were spun at 1000 rpm and washed with lysis buffer. Loading dye, as in western blotting, was added and denatured. Ran on gel with input and probed with corresponding antibody.

2.3 RNA analysis

2.3.1 DNA/RNA extraction

DNA and RNA was extracted from cell pellets using an All Prep RNA/DNA kit (Qiagen) or RNA singly using the RNeasy mini Prep kit (Qiagen). Quantities were measured using a Nanodrop spectrophotometer 1000 (ThermoFisher). RNA was converted to cDNA using a superscript cDNA VILO kit (Invitrogen).

2.3.2 Real-time Quantitative PCR

mRNA levels were measured using qPCR. DyNAmo Flash SYBR Green (ThermoFisher) mastermix was used and the RotorGene 6000 (Corbett Life Sciences/Qiagen). Standards were made from a mix of cDNAs and serially diluted to 1:1000. Primers were designed so that they would traverse two exons so as to select from cDNA rather than gDNA is case of contamination. All standards and samples

were performed in duplicate. Each sample was standardised to housekeeping genes that were tested in optimisation experiments to show they were unaffected by experimental conditions.

2.4 DNA analysis

2.4.1 Sanger Sequencing

Sequencing was performed at the Human Genetics Unit, IGMM. Primers were supplies at 100nM stocks.

2.4.1.1 PCR

Amplification PCRs were performed with 1ng DNA starting material, 0.4nM primers for region 1 and 0.5ng DNA, 0.2nM for region 2. GoTaq Green Master Mix (Promega) was used and each PCR reaction contained a template free control. The PCR programme included at 4 minute 10 second extension period.

Amplification Primers	Sequence
Region 1 Forward	GGATCACGAGGTCAGGAGTT
Region 1 Reverse	TTTGTGCGGTGGGAATTCTC
Region 2 Forward	acgattctgaggcaaaggga
Region 2 Reverse	cgtagggaatggagaggtgg

2.4.1.2 Sequence analysis:

An R program (Clustal) was used to compile the individual sequence from each sequencing primer into one contig for each patient. This program trimmed the ends of the sequence data of bad quality sequence and then aligned the sequence to a reference sequence (reference sequence was taken from UCSC). Deviations from the reference sequence were marked with an *. The chromatograms were checked to determine if they were of poor quality and a different call should be made, or if they were genuine SNPs.

2.4.2 Pyrosequencing

DNA was first bilsuflite converted using the EZ-96 DNA methylation Kit (Zymo Research). Pyrosequencing PCR primers were designed using the Q24 PyroMark software and were designed to avoid covering a CpG site so as to avoid introducing bias into the amplification. Either the forward primer or the reverse primer was biotinylated. Sequencing primers were also designed using the Q24 PyroMark program. Assays were also designed using this software for each area of interest. This determines sequence of nucleotides to be added to the reaction. Each sample was performed in duplicate and there was a no template control in each run. Samples were only included for analysis if they were the highest quality (denoted 'blue' by the software) and an average of the two readings taken.

2.5 Buffers

Lysis buffer	0.5% NP40 50nM Tris pH 7-8 150nM NaCl dH2O to volume
Transfer buffer	25mM TRis 190mM glycerine 20% methanol
Running buffer	0.125M Tris-HCl

3 Results

3.1 Sequencing

3.1.1 Introduction

A 6.5kb region encompassing the mir21 primary transcript was sequenced in 300 people. The aim was to discover novel SNPs that may be in these Scottish patients, as well as investigating whether this correlates with disease state and/or methylation of the region.

SNPs are single point mutations. Comparing SNPs found in cases compared to controls in association studies is a technique used to find areas of the genome that may be relevant to disease. The SNPs themselves maybe causative, or they may tag an area that they are in linkage disequilibrium with, which may have functional element in it. A SNP may be causative in several ways- if it is in a protein coding region the change may alter the amino acid sequence, changing the protein structure. If it is not in a protein coding region it may still be in the transcript for microRNAs, affect splice sites or binding sites for transcription factors and microRNAs.

If a SNP is found an R program, SuRFR, can be used to prioritise SNPs found in order of likelihood of their being functional (Ryan et al. 2014). It uses known functional annotation data about the region -conservation scores, DNase hypersensitivity, chromatin states and transcription factor binding sites- to prioritise the found SNPs so further investigation can be focussed on the most likely to be functional.

There are two SNPs nearby already associated with IBD- rs1292053, identified by a large meta-analysis of GWAS studies by Jostins et al (Jostins et al. 2012) and rs8077981 (Raelson et al. 2007), identified from a smaller cohort of Quebeci people. The first SNP is 44.9kb upstream from mir21 in the TUBD1 gene and the second 90.1kb further downstream in the BCAS3 gene. There has also been an meQTL identified in this region. Two SNPs, rs10853015 and rs8078424, both correlate strongly with methylation at the VMP1 locus. They are also in linkage disequilibrium with rs1292053, the initial SNP identified by Jostins et al. These findings together indicate this region is of particular interest to IBD.

SNPs found in the region sequenced, which encompasses the miR-r21 primary transcript could be interesting for various reasons, despite not being wholly in a coding region. They may alter expression of miR-21, they may affect the binding of transcription factors or they may alter the sequence of the mature miR-21 transcript which would affect its ability to its targets.

3.1.2 Aim

The aim of this experiment is to sequence this area in a cohort of 300 patients. Any SNPs identified can be compared to known SNPs to see if any are novel. They can be compared to the disease state of the individuals to establish if they correlate with disease and also compared to methylation data from the region which has been established by research previously undertaken in the lab an expression levels of VMP1 and miR-21.

3.1.3 Methods

These were 100 CD patients, 100 UC patients and 100 HC. This group was selected from a large database of people recruited at the Western General Hospital in Edinburgh. They have been genotyped at the strongest SNP associated with disease-rs59886127. Healthy controls were mostly lab volunteers or suspected IBD patients that actually did not have IBD but other gastrointestinal complaints, such as IBS. These may be an interesting subgroup that may help distinguish between symptomatic healthy people with actual IBD patients. This will be powered enough to identify SNPs in 1% of cases.

Diagnosis	Female	Male
HC	49.1%	50.9%
UC	59.6%	40.4%
CD	50.4%	49.6%

Diagnosis	Current	Ex	Never
HC	19.2%	33.3%	47.5%
UC	43.8%	22.3%	33.9%
CD	10.6%	38.5%	50.9%

Breakdown of diagnosis by smoking status.

Diagnosis	Median Age	Interquartile range
HC	34.23836	18.86849
UC	34.92603	22.72192
CD	31.93425	25.33562

Breakdown of diagnosis by age

Area selected for sequencing: The area chosen to be sequenced was about 6.5kb which encompasses the sequence for the primary transcript for miR-21.

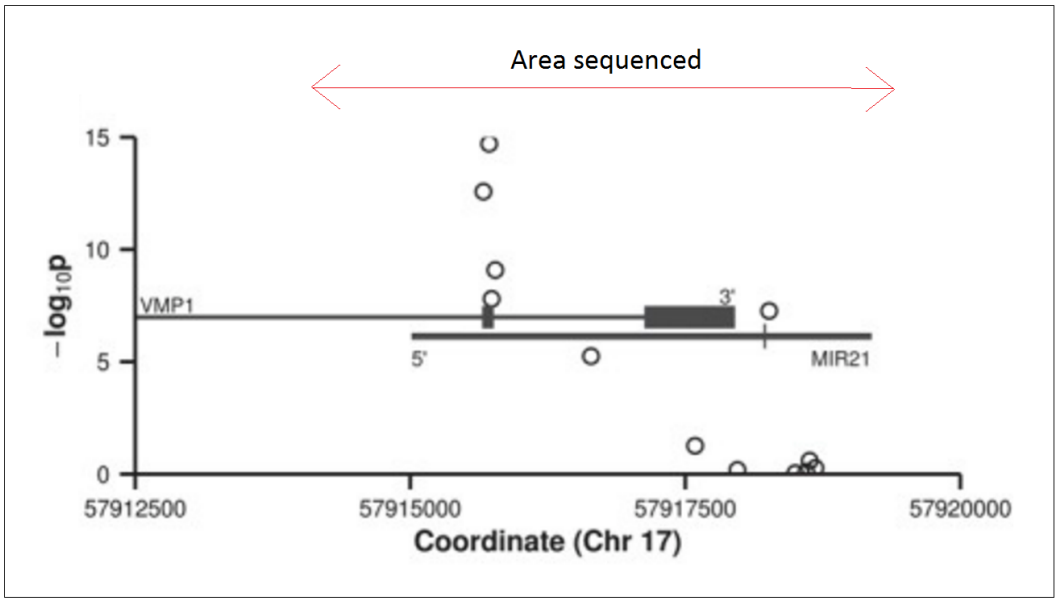

Figure 5: Schematic adapted from Adams et al showing the primary transcript (line) and mature transcript (bar) in respect to VMP1. Red line indicates area sequenced.

This area at the 3' end of the VMP1 gene and the area sequenced will incorporate the final exon of VMP1.

Data analysis

The data was outputted into an ab1 file. This was analysed using R- the sequence was quality controlled by trimming the ends until they had 10 bases in a row with no 'N' base calls. They were aligned to a reference sequence from UCSC using a program called Clustal Omega (Goujon et al. 2010)). The sequence quality was scored by the proportion of ATCG nucleotide calls and the proportion of sequence that aligns to any other sequence data. Finally, the program noted any base that was different from the sequence and these were manually checked to identify if they were genuine SNPs, heterozygous bases or poor quality sequence that could be ignored.

3.1.3 Results

200 of the 300 patients have been sequenced so far. Sequencing data of the remaining 100 is still ongoing. The sequence data has been quality controlled and aligned to the reference sequence. Sequence primers were optimised on test patients and only those chosen have been used in the remaining patients. However, some primers have still given data of variable quality. As such some parts are lacking sequencing data. 9 patients have very poor quality sequence data (more than 2 standard deviations from the median), possibly because the original amplification PCRs failed to work, and so have been excluded from analysis.

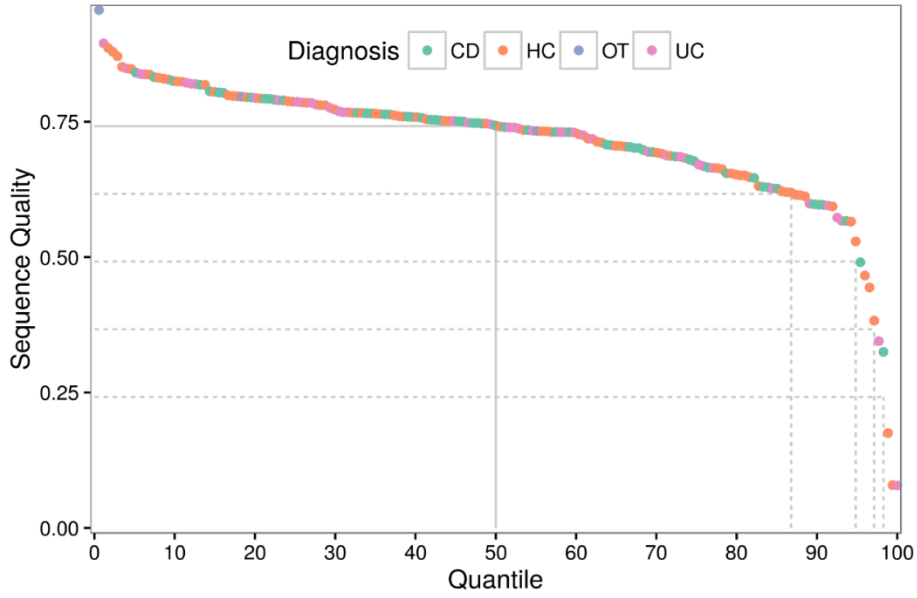

Figure 6: Graph showing quality of sequence for each patient. Patients 2 standard deviations from the median quality were excluded. Each dotted line represents 1 standard deviation from the median.

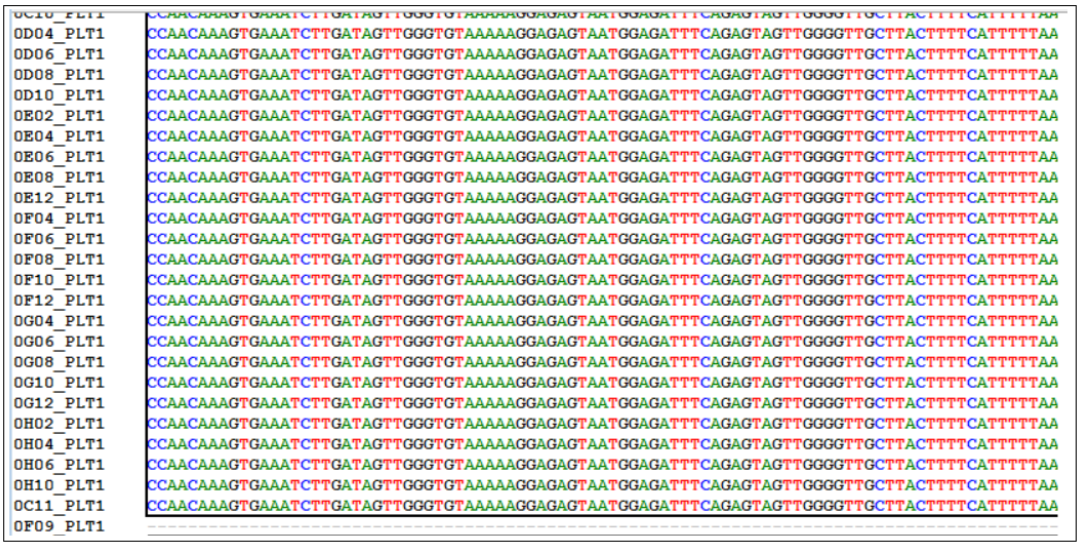

Figure 7: Example of aligned sequence data that has been quality controlled. No data for final patient

In the data sequenced and analysed so far there have been no SNPs identified.

3.1.4 Conclusions

So far 200 of 300 patients have been sequenced – although parts of this sequence need to be improved upon. No SNPs have been discovered so far, indicating this region is highly conserved.

3.2 Pyrosequencing

3.2.1 Introduction

Currently faecal calprotectin tests are used clinically as a non-invasive way to distinguish between IBD and non-IBD. However, a blood test may be more sensitive, or specific and will be easier to perform than a faecal test. Tests that require a faecal sample also have issues with compliance from patients and are expensive. Further, a blood test that could distinguish IBD from non-IBD or even UC from CD could help significantly with earlier, easier, more accurate diagnosis and may prevent unnecessary endoscopies. The potential for DNA methylation to be used as a biomarker test has previously been explored in cancer (Lofton-Day et al. 2008), and for IBD in paediatric cases (Adams et al. 2014).

3.2.2 Aim

The aim for these experiments was to develop a model for diagnosing IBD from HC using the methylation status at particular CpG sites by pyrosequencing. This should then be validated in a larger cohort.

3.2.3 Methods

The initial stage use tested 8 CpG sites in 20 HC, 20 UC and 20 CD patients by pyrosequencing. These 8 sites were chosen by using epigenome-wide methylation data from a previous study undertaken by the lab (Ventham 2016). This was a study involving 431 adult patients and a 450K methylation chip. This identified 439 CpG sites that were differently methylated in IBD patients compared with healthy controls.

Two of the probes were chosen to identify healthy controls from IBD patients. An R package, MASS, (Venables, W. N. & Ripley 2002) was used to perform a linear discriminant analysis using a combination of 2 CpG sites from the top 200 significant probes. Using more than 2 CpG sites did not improve sensitivity and specificity significantly. The top model involved CpG sites in the genes Ribosomal Protein S6 Kinase S2 (RPS6KA2) and LSP1. These were able to distinguish between IBD and HC with a specificity of 82.1% of and sensitivity of 76.9%.

The remaining 6 probes were identified in an attempt to distinguish CD from UC. For this only the people the original analysis had selected as being IBD (regardless of whether they were or not) were used and a further linear discriminant analysis using combinations of 3 CpG sites from the top 20 for differences between UC and CD in the 450k data. 6 unique CpG sites produced the top 3 models. This approach was justified as it would mimic the approach taken if the specific disease state of the patients were unknown at time of analysing. In total there were 8 CpG sites chosen to be pyrosequenced, these were:

CpG reference	Gene
cg17501210	RPS6KA2
cg09569850	Chr2: 64955034
cg02573091	ANKDD
cg01005605	BCL11A
cg04666911	LSP1
cg05304729	MNDA
cg25069807	TK1
cg05329352	ADRA2

Training set

A smaller cohort of 20 CD patients vs 20 UC patients vs 20 HC people were chosen to be analysed by pyrosequencing and then tested by linear discriminant analysis. These patients were selected from the original 300 and were closely matched with age, sex and smoking status.

Validation

The top linear model (chosen by fitting the models to the same data, and choosing the model that works best) should then be tested in the remaining patients from the 300 cohort.

3.2.4 Results

Results were only included if the data was deemed to be top quality by the pyrosequencer software. Each sample was pyrosequenced in duplicate and an average percentage methylation taken for each CpG.

Methylation differences measured using Mann-Whitney U tests as previous studies have shown the variance in methylation in HC vs IBD to be very different.

3.2.4.1 RPS6KA2

RPS6KA2 was one of the top findings the 450k methylation chip study undertaken by Adams et al., and was replicated using pyrosequencing in adults (P-value of $4.4x10^{-5}$. It was also the top finding (uncorrected P-value of $2.7x10^{-22}$), for the second 450K study undertaken in the lab, in 431 adults. RPS6KA2 is a ribosomal S6 kinase and genetic variants in RPS6KA2 have been associated with colon cancer (Ding et al. 2013). Three sequential CpGs were sequenced in the same assay in healthy controls, Crohn's disease patients and ulcerative colitis patients using pyrosequencing.

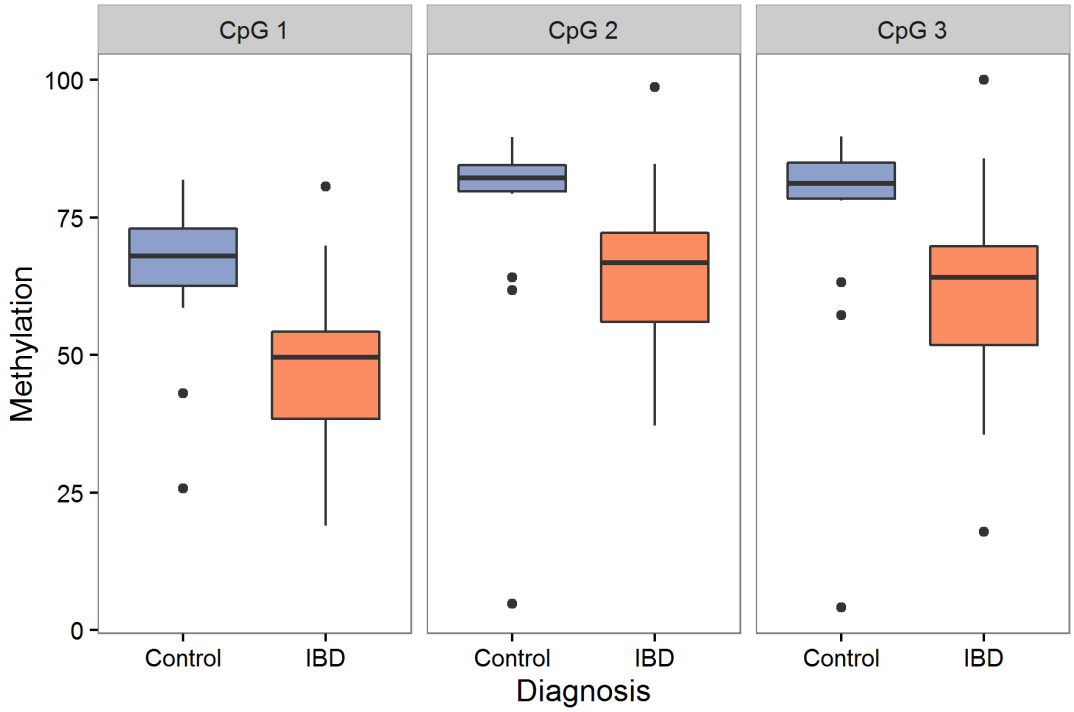

Figure 8: Boxplot showing methylation at three CpG sites in the RPS6KA2 gene. Differences between healthy controls and IBD patients are statistically significant. P- values; CpG 1: 1.90408×10^{-5}, CpG 2: 0.0001307362, CpG 3: 9.977015×10^{-5}. N=55, HC=16, IBD=39

These data replicate the original finding; there is a significant difference in hypomethylation in the RPS6KA2 gene in healthy controls compared to IBD patients. There is a median difference of 18.44% between HC and IBD and there is a greater variance in methylation in IBD compared to HC.

The IBD data can be further broken down into CD and UC.

CpG 1 in RPS6KA2

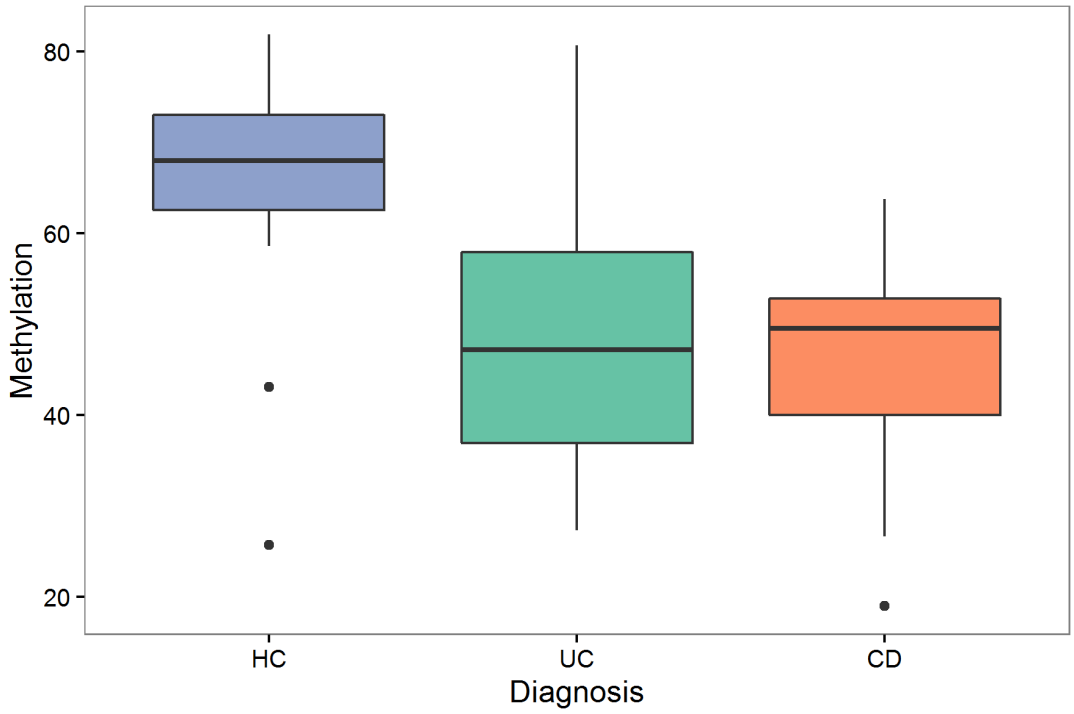

Figure 9: Boxplot showing the methylation at CpG 1 in RPS6KA2 between HC, UC and CD. N=55, HC=16, UC=21, CD=18. Significant difference between HC and CD (p-value 4.4 x10^{-5}), and HC and UC (p-value 0.0006916).

CpG 1 gave the strongest finding, and so has been shown here. There is a statistically significant difference between HC and both UC and CD, although no difference between UC and CD.

3.2.4.2 BCL11A

B-cell lymphoma/leukemia 11A (BCL11A) is a transcriptional regulator that is involved in breast cancer (Khaled et al. 2015), and its downregulation has been shown to induce apoptosis (Gao et al. 2013). Two sequential CpG sites were pyrosequenced in BCL11A.

CpG 1 in BCL11A

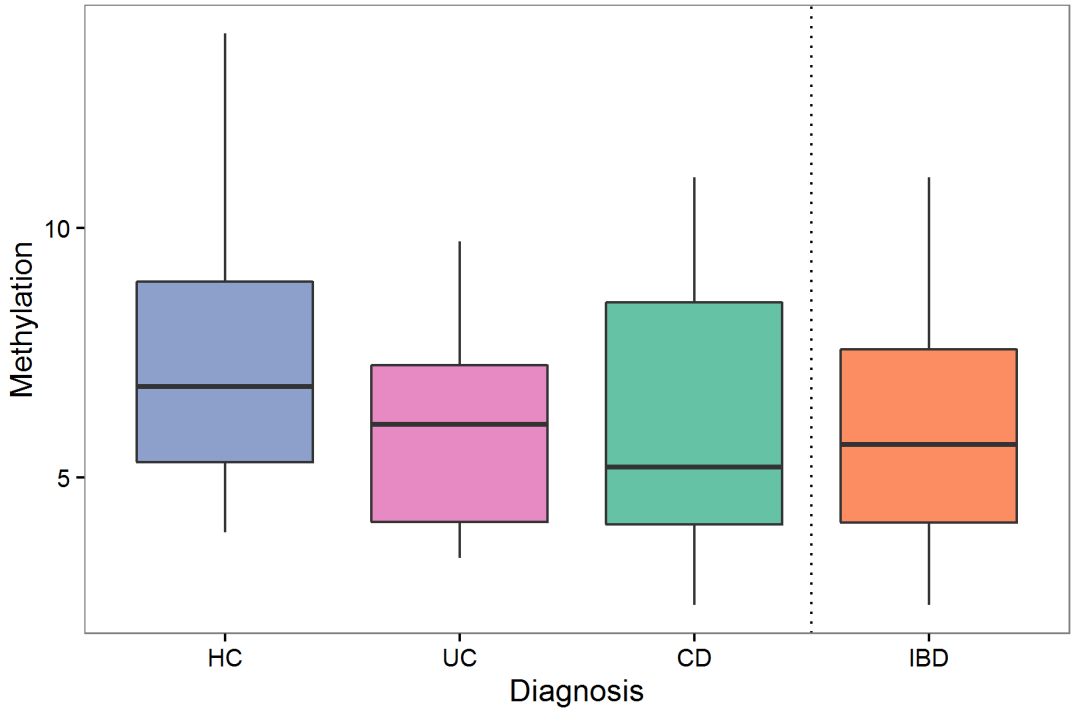

Figure 10: Methylation at BCL11A distinguished by diagnosis for IBD. No significant differences between HC and IBD diagnoses. IBD classs (on RHS) is composed of UC and CD, also shown here. N=54, HC=18, UC=19, CD=17

There is no significant difference between healthy controls and any of the diagnoses.

CpG 2 in BCL11A

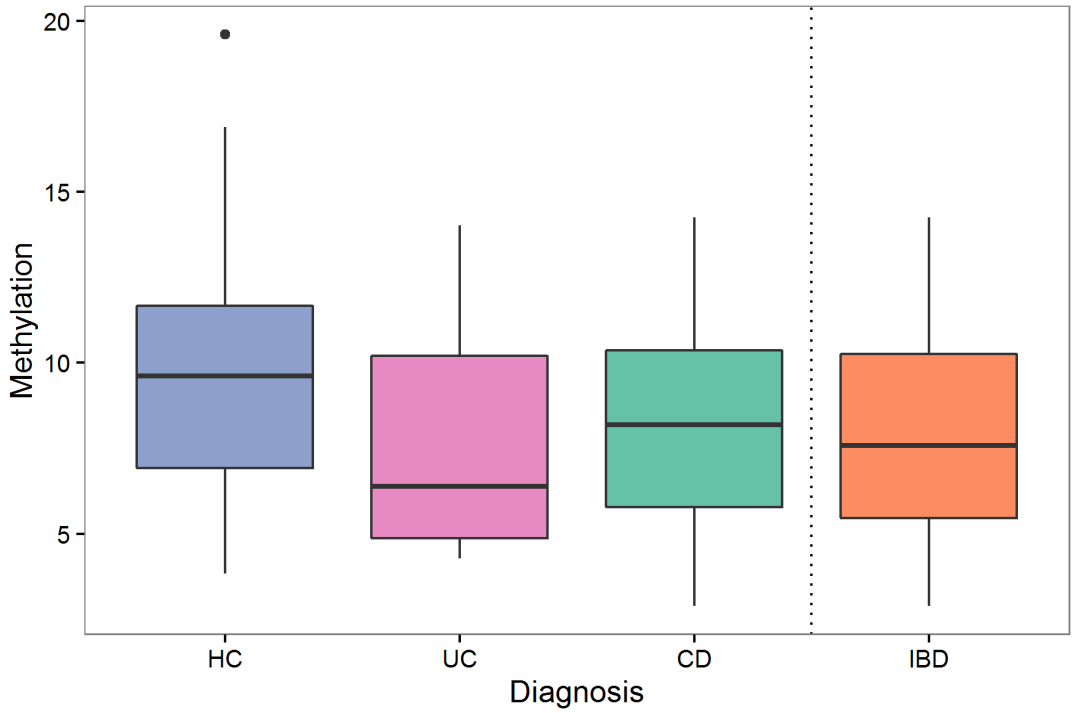

Figure 11: Methylation at the second CpG in BCL11A distinguished by diagnosis for IBD. No significant differences between HC and IBD diagnoses. N=54, HC=18, UC=19, CD=17

There is no significant difference between any of the classes for the second CpG at BCL11A, however, there is a difference in median methylation of 3.23% between HC and UC, and 1.43% in HC and CD.

3.2.4.3 Chr2: 64955034

CpG 1 in Chr2: 64955034

Two sequential CpG sites were pyrosequenced in HC and IBD patients. This probe was not in a gene and so left named as the location of the 450K methylation probe in chromosome 2.

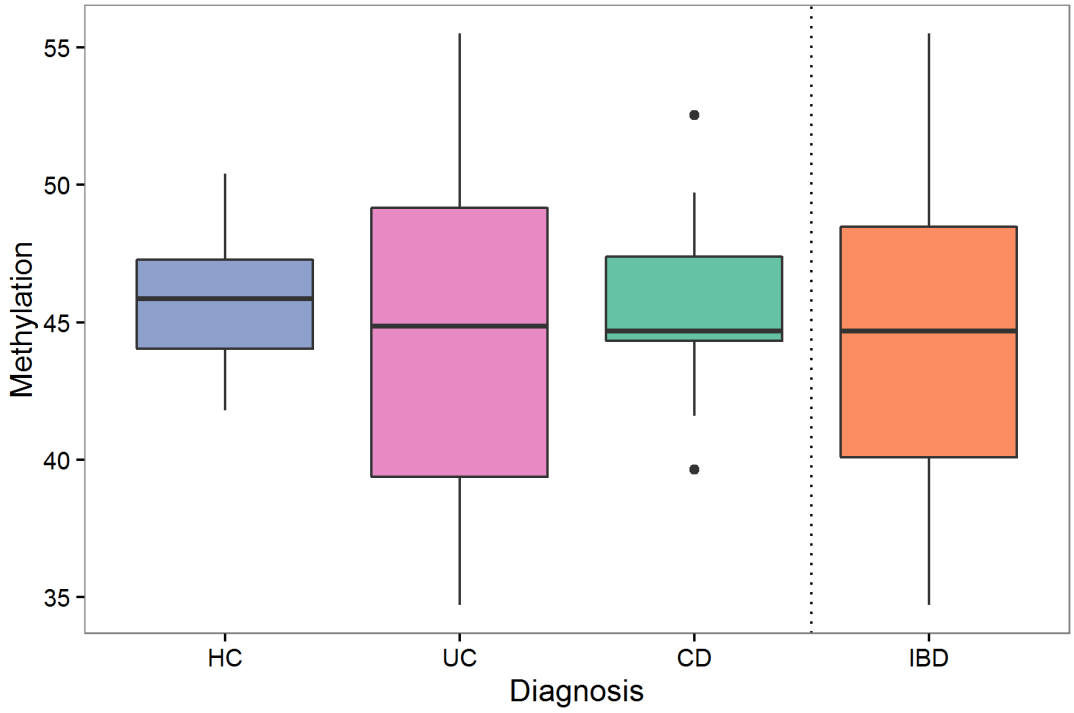

Figure 12: Boxplot to show methylation at first CpG in Chr2: 64955034. No significant differences between HC and IBD diagnoses. HC=7, UC=12, CD=9.

There is no significant difference in methylation at this CpG between any of the classes, however there is a large increase in variance between HC and UC patients.

CpG 2 in Chr2: 64955034

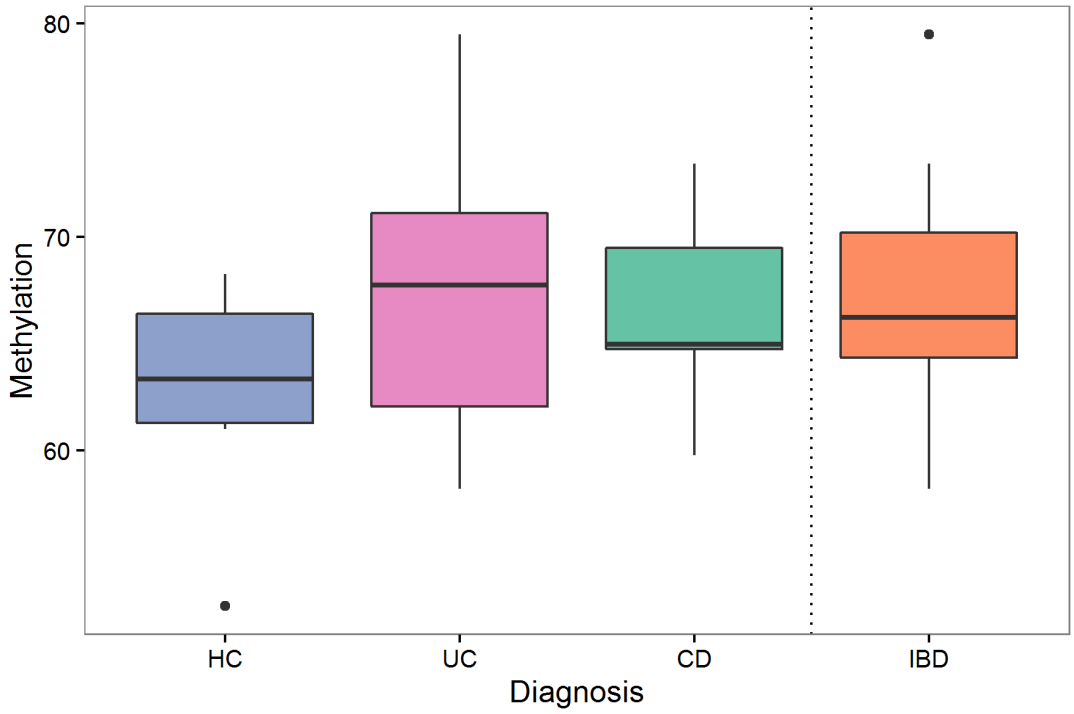

Figure 13: Boxplot to show methylation at second CpG in Chr2: 64955034. No significant differences between HC and IBD diagnoses. HC=7, UC=12, CD=9.

Again there was no significant difference between the classes, but an increase in variance in the UC patients.

3.4.3.4 LSP1

Lymphocyte-Specific Protein 1 (LSP1) is a cytoskeletal protein that is predominately expressed in leukocytes (Jongstra-Bilen et al. 1992). Two sequential CpGs were analysed by pyrosequencing.

CpG 1 in LSP1

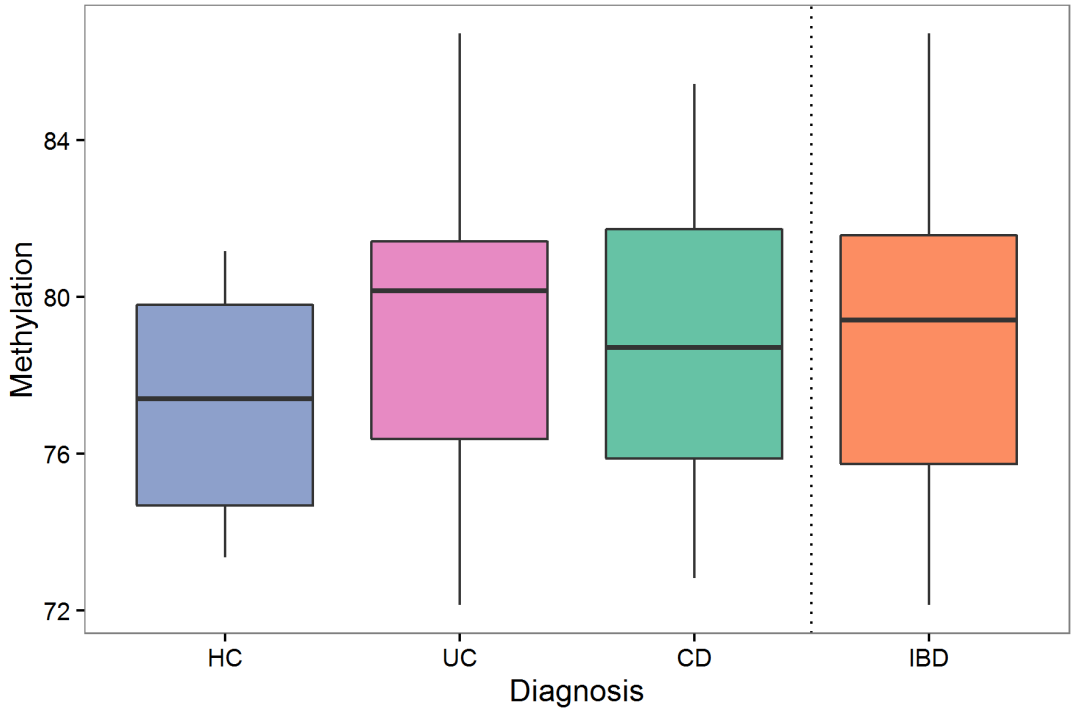

Figure 14: Boxplot to show LSP1 CpG 1 methylation in HC vs IBD diagnoses. N=54, HC=18, UC=19 and CD=17

There was no significant difference in methylation between the diagnoses, although the median UC value is hypermethylated by 2.8% compared to HC.

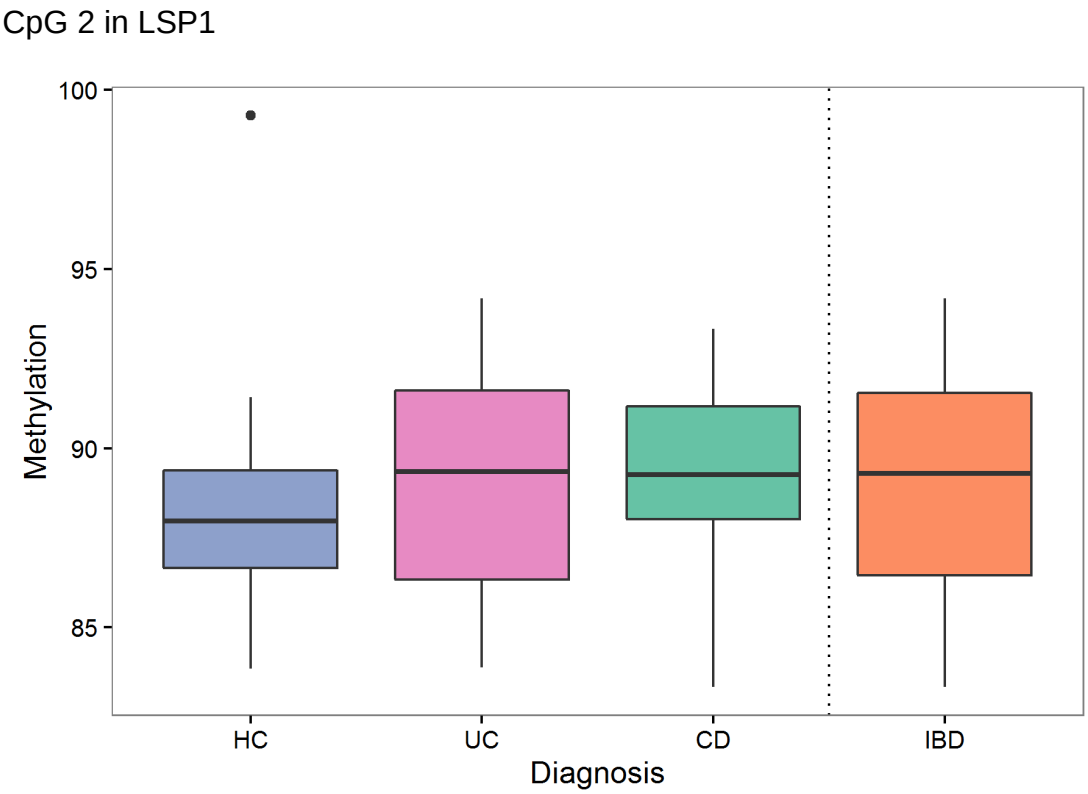

Figure 15: Boxplot to show methylation at second CpG site in LSP1. N=54. HC=18, UC=19 and CD=17

The second CpG in this assay also does not show a significant difference in methylation between diagnoses, however the IBD median is hypermethlyated by 1.3% compared to controls.

3.2.4.5 MNDA

MNDA (Myleiod cell nuclear differentiation antigen) is expressed only in monocytes and in response to interferons. It is involved in cell differentiation and apoptosis (Metcalf et al. 2014).

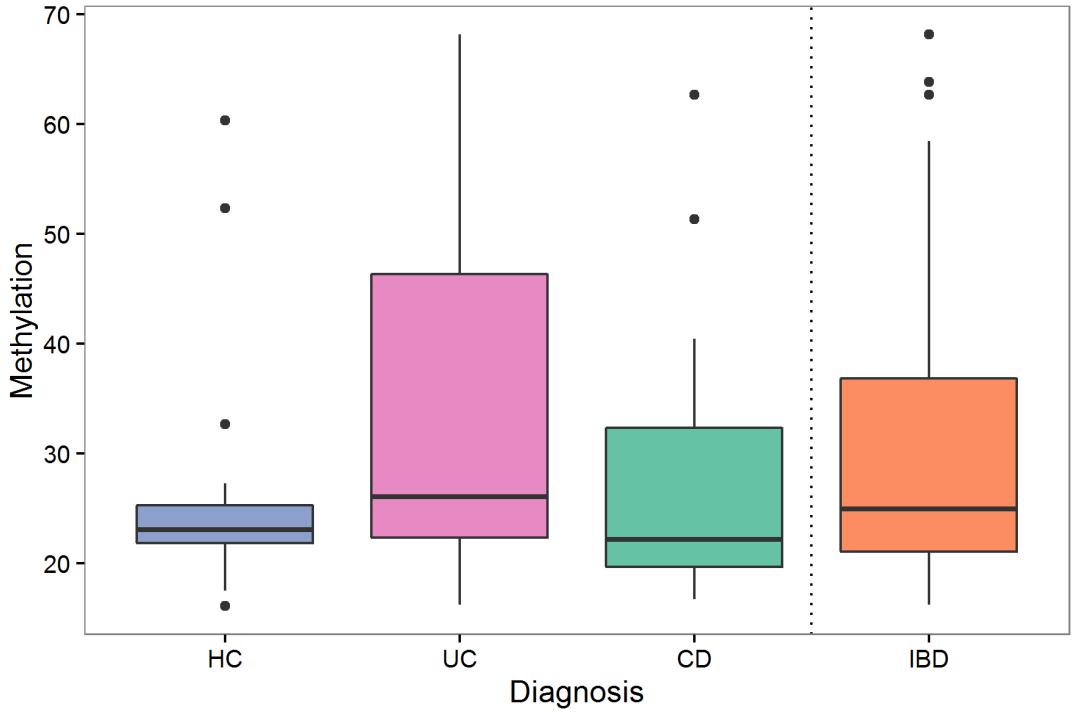

Figure 16: Boxplot showing methylation in MNDA for HC and IBD diagnoses. N=54, HC=18, UC=19 and CD=17

The MNDA CpG does not show significant differences between diagnoses, although there is a large increase (25.9%) in variance in UC compared to HC.

3.2.4.6 TK1

Thymidine Kinase 1(TK1) is a thymidine kinase. Two sequential CpG sites in TK1 were pyrosequenced.

CpG 1 in TK1

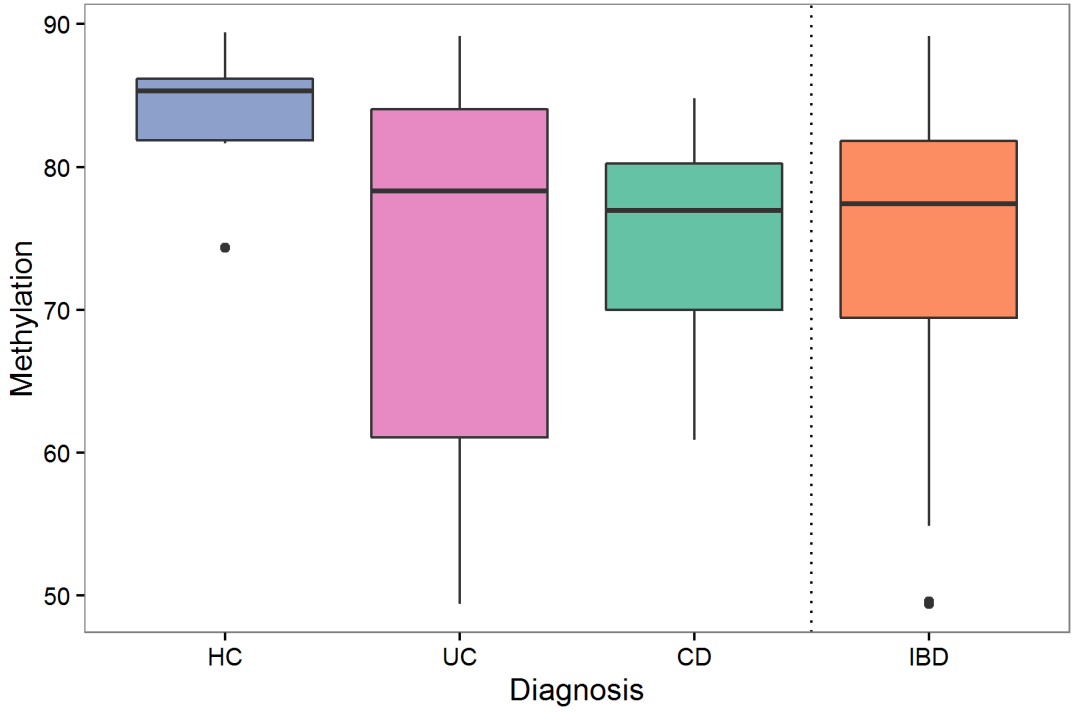

Figure 17: Boxplot showing hypomethylation at CpG1 in TK1 between HC and IBD. N=36, HC=10, UC=14, CD=12.

There is a statistically significant difference between HC and IBD patients for the first CpG site (P- value 0.003776). There is a significant difference between HC and both UC (P value= 0.04172) and CD (P- value 0.0091935), and no difference between CD and UC.

CpG 2 in TK1

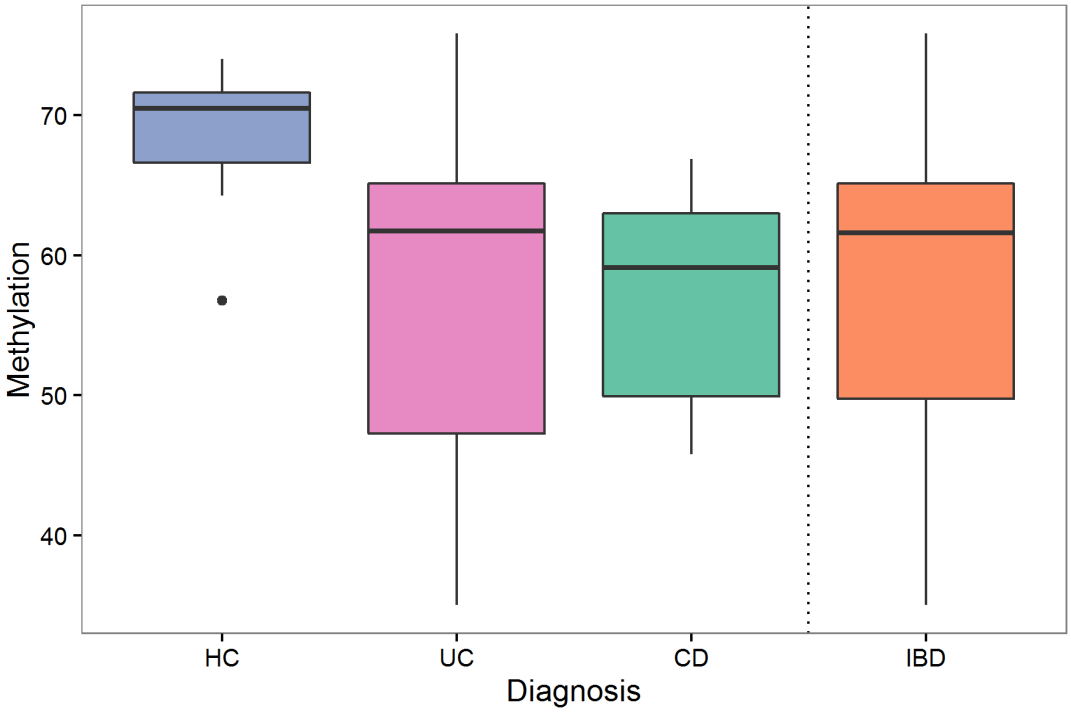

Figure 18: Boxplot showing methylation at CpG 2 in TK1.

There is a statistically significant difference between HC and IBD patients for the second CpG also (P- value 0.001308). There is a significant difference between HC and both UC (P value= 0.01179) and CD (P- value 0.000832), and no difference between CD and UC. The differences in the second CpG measured is stronger between HC and IBD than the first CpG measured.

3.2.4.7 ANKDD1A and ADRA2AC

The two CpG sites in these genes failed to give good enough quality data to be sequenced. This can happen in pyrosequencing if the CpG site of interest is in a CpG island- the sequencing primer needs to be able to bind to sequence free from CpGs itself so as not to bias results. Also, if the area is highly repetitive it is difficult to get good quality data. As such, neither of these sites would be suitable for an easy diagnostic test by pyrosequencing.

3.2.4.8 VMP1

The significant CpG site in VMP1 has previously been pyrosequenced by the lab in many of the patients in this cohort. The results have been included here, to demonstrate the difference in HC and IBD.

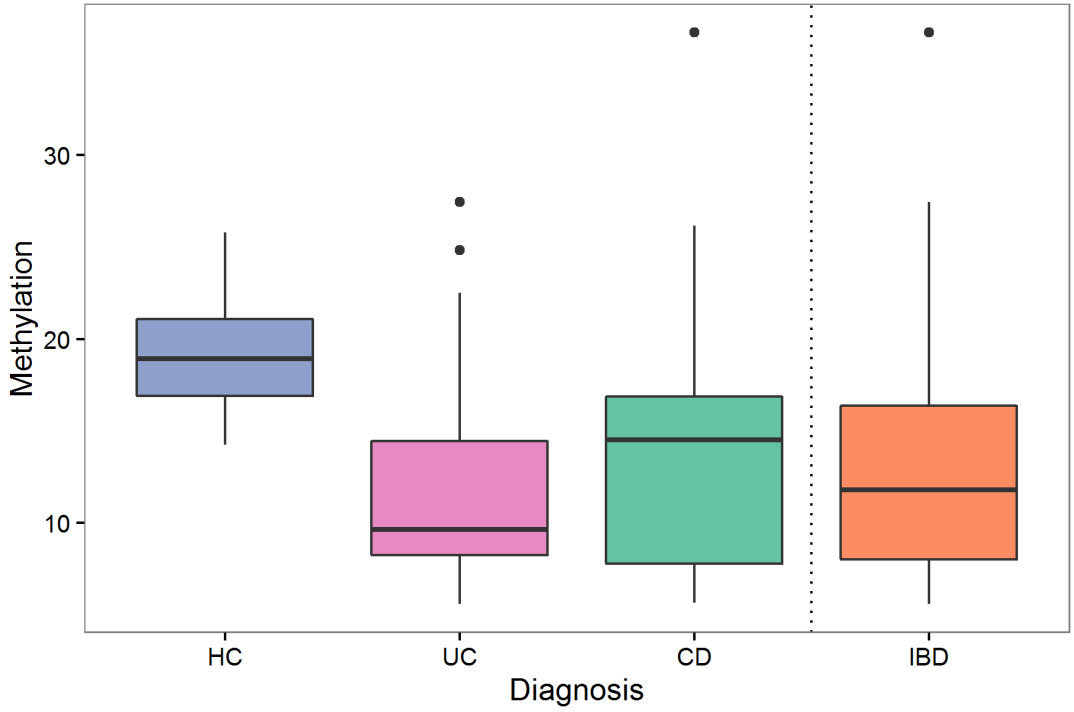

Figure 19: Boxplot to show the methylation at VMP1 in HC and IBD patients, previously determined by the lab with pyrosequencing. N=38, HC=8, UC=17, CD=13

There is a significant difference between HC and UC (p- value 0.0156), and no significant difference between HC and CD or UC and CD.

3.2.5 Linear discriminant analysis

Linear discriminant analysis uses a set of observations from a training set in order to make a prediction in another set. For this data the methylation status at two or more CpG sites will be used to make a model that will attempt to distinguish IBD from HC, and maybe UC from CD.

Not every patient had complete data for every CpG site which is required to develop a linear discriminant analysis model. For four CpG sites- RPS6KA2, LSP1, BCL11A and MNDA there were 49 patients with complete data.
The breakdown of these patients is;

	CD	HC	UC
Female	8	8	11
Male	8	7	7

This model was able to predict IBD with a sensitivity of 91.2% and a specificity of 66.7%. Using the same probes When only the IBD patients were surveyed it was able to identify CD with a sensitivity of 75% and UC with a sensitivity of 55.6%.

33 Patients had complete data for 5 probes; the above and TK1.

	CD	HC	UC
Female	4	4	6
Male	7	6	6

This model was able to identify IBD with a sensitivity of 91.7%, and a specificity of 80%. It was able to distinguish between CD with a sensitivity of 75% and UC with a sensitivity of 56.7%.

3.2.6 Conclusions

The model was able to identify IBD with a high sensitivity. However, the specificity is not very great and it is poor at distinguishing between UC and CD. Some repeats of the data are required before the full linear discriminant analysis can be performed on all CpG probes which will give greater power for developing the model. When the top probes have been chosen it will be validated in the larger cohort of 300 patients.

3.3 Mir21 and autophagy

3.3.1 Introduction

There is some evidence already that mir21 is involved in autophagy, but there is no data on whether mir21 affects VMP1-its host gene and key regulator of autophagy. Mir21 knockdown and mimics have previously been used to alter mir21 expression and investigate the effects on other genes, such as Programmed Cell Death 4 (PDCD4). Autophagy can be measured by LC3I-LC3II conversion by western blot. By altering mir21 expression and measuring VMP1 and LC3 expression it will be possible to ascertain the effects of mir21 on autophagy, and if this is being affected through VMP1.

3.3.2 Aims

The aims of these experiments was to get a working assay for autophagy and a western antibody for VMP1, to successfully knockdown and overexpress mir21 in cells and measure the effect on VMP1, and to establish an interaction between VMP1 and NOD2.

3.3.2 Optimisation

Antibodies

An antibody for VMP1 was optimised for western blotting. The antibody was tested at concentrations of 1:200- 1:2000 and on both PVDF and nitrocellulose membrane. Membranes were blocked and probed with either milk/PBS-T solution or BSA/PBS-T solution. The blots failed to produce a signal at all. Finally, alternating blocking with milk/PBS-T and probing in BSA/PBS-T or vice versa was tried and this produced a signal. The final conditions for successful western blotting of VMP1 were a concentration of 1:1000 antibody and PVDF membrane. The membrane was blocked for 1 hour at room temperature in a 5% BSA/PBS-T solution, washed with PBS-T, probed with VMP1 primary antibody overnight in a 5% milk/PBS-T at 4degrees, washed again with PBS-T and then finally probed with rabbit secondary antibody in a 5% milk/PBS-T solution for 1 hour at room temperature.

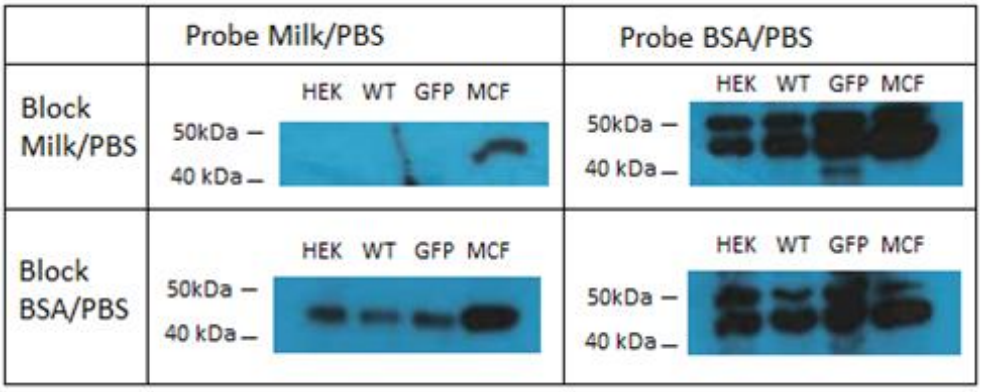

Figure 20: Optimisation western blots. HEK=HEK293T cells. WT=NOD2 WT overexpressing HEK cells. GFP= GFP-LC3 overexpressing HEK cells. MCF=MCF-7 cells

A western blot was also optimised for the LC3 antibody: the membrane was blocked and probed in a milk solution, and the final conditions were; a concentration of 1:1000, 5% milk/PBS solution.

Autophagy Assay

HEK-GFP-LC3 and THP1 cells were stimulated with bafilomycin and starved to induce autophagy. Bafilomycin stimulation was at 20uM concentration and for 3 or 6 hours. 3 hours gave as good results as 6 hours so was used for final experiments. Cells were starved using a starvation media for the same amount of time. Measuring the ratio of LC3-I to LC3-II by western blot indicates the level of autophagy. Bafilomycin prevents the maturation of autophagosomes, causing the accumulation of LC3-II

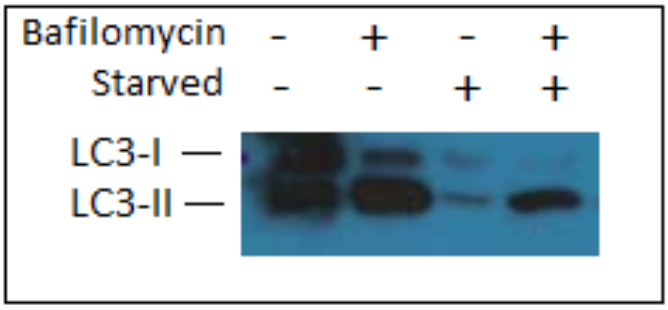

Figure 51: Autophagy assay in THP-1 cells, using LC3 antibody

This blot demonstrates that when starved and treated with bafilomycin there is a high ratio of LC3-II to LC3-I. When the cells are treated with bafilomycin but not starved

(autophagy has not been induced) there is an accumulation of LC3-II, but it is not significantly higher than the level of LC3-I. When the cells were starved but not treated with bafilomycin the levels of LC3 are similar, but the bands are weak. This is a loading error.

HEK293 cells that are stably transfected with a GFP-LC3 construct can also be imaged to show the formation of autophagosomes in the cells. This experiment is shown here with rapamycin stimulation. Rapamycin inhibits mTORC1, an inhibitor of autophagy, causing autophagy to be activated. Upon autophagy activated autophagosomes are formed, and GFP-LC3 form bright punctae that can be imaged. However, this method is not a quantitative measure of autophagy and can be subjective.

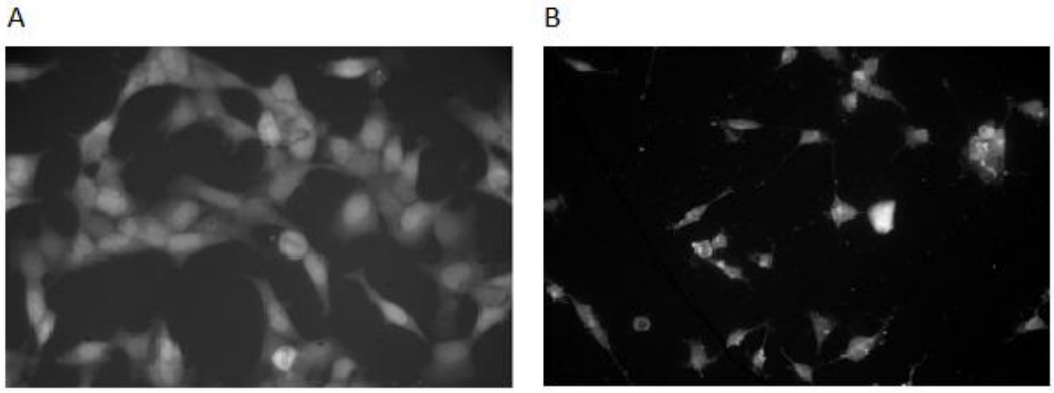

Figure 22: HEK293 cells stably transfected with GFP-LC3 construct. A= Control cells grown in normal media for 24hr. B= Cells starved and treated with rapamycin 50nM for 24hr

Mir21 knockdown/ overexpression

This was performed in THP1 cells, and PDCD4 mRNA levels were used as an indicator of whether the knockdown or over expression was successful. Conditions tested were 100nM, 75nM, 50nM for the inhibitor or 10nM, 7.5nM or 5nM for the mimic. Time points of 24hr, 48hr and 72 hrs were used.

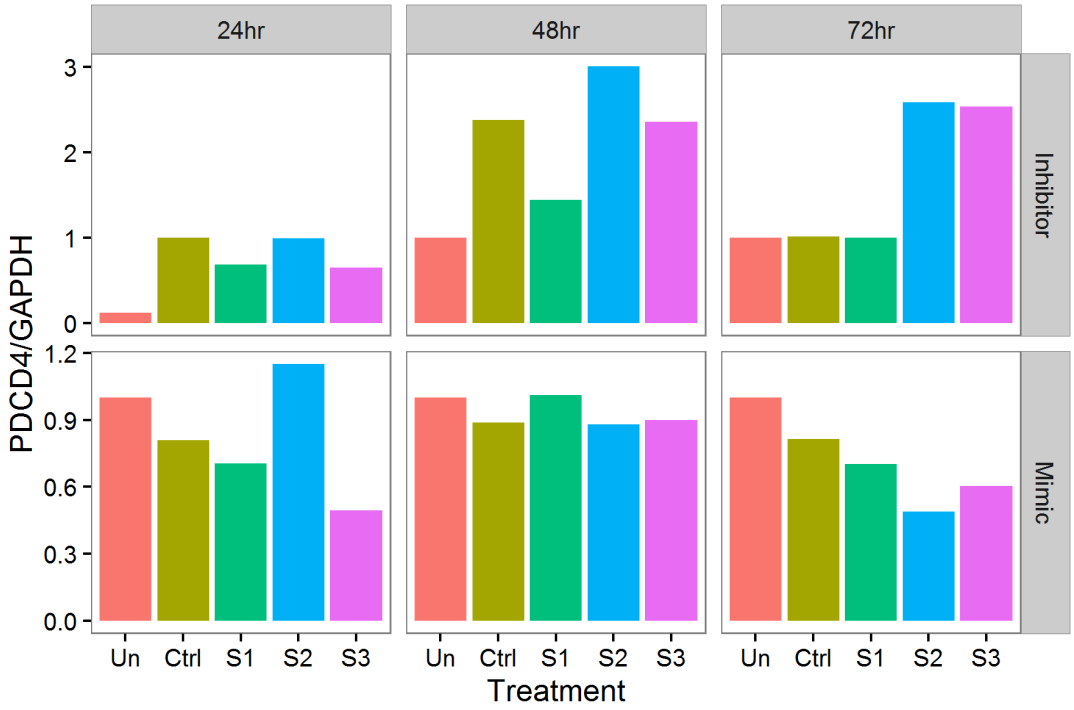

Figure 23: Relative PDCD4 expression to GAPDH. Un= Untransfected, Ctrl= Control siRNA, S1= 50nM Inhibitor or 5nM Mimic, S2= 75nM inhibitor or 7.5nM mimic, S3= 100nM inhibitor or 10nM mimic.

These experiments showed a decrease of nearly 50% of PDCD4 when cells were treated for miR-21mimic for 72hr at 7.5nM and a 2.5X times increase of PDCD4 when treated with 75nM inhibitor for 72hrs. These conditions were chosen therefore for further experiments.

3.3.4 Co-immunoprecipitation of VMP1 and HA-NOD2

Co-immunoprecipitation experiments were performed on HEK cells that are stably transfected with HA-NOD2 constructs. VMP1 antibody was conjugated to protein G agarose beads and incubated with samples. There is not yet a working antibody for western blotting NOD2 and so cells lines that have been stably transfected with a HA-NOD2 construct were used, and a HA antibody was used to probe the lysate.

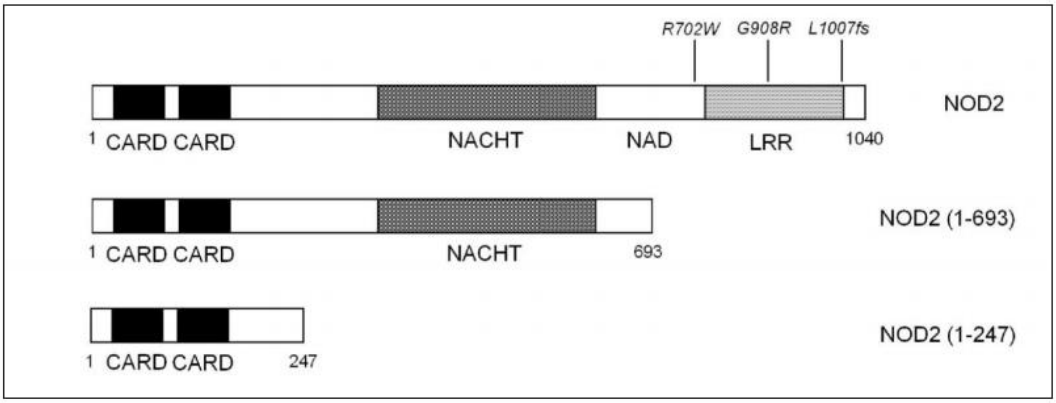

Figure24: Schematic adapted from Stevens et al demonstrating NOD 2 WT, and NOD2 deletion mutants

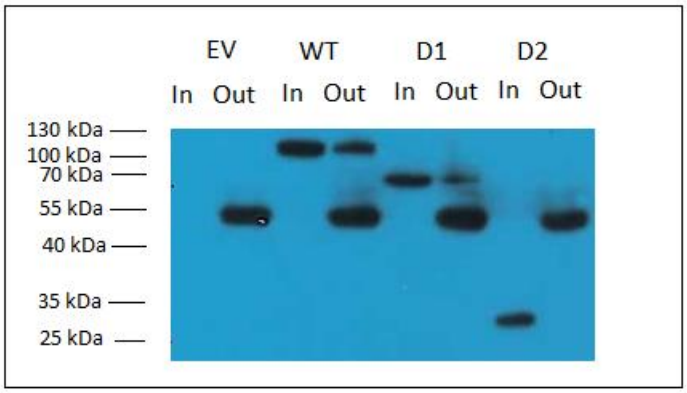

Figure25: Western blot using HA antibody. EV= HEK293 cells transfected with an empty vector. WT= HEK293 cells transfected with NOD2 wild type. D1= HEK293 cells stably transfected with partially deleted NOD2 construct, 1-693, D2= HEK293 cells stably transfected with partially deleted NOD2 construct, 1-247

VMP1 was successfully immunoprecipitated with HA-NOD2. This demonstrates that the NOD2 interact with VMP1. This interaction was disrupted in the second NOD2 deletion, indicating the NACHT domain is crucial for this interaction. The reverse co-immunoprecipitation experiment- IP using VMP1 and western blot with HA- did not work successfully.

3.3.5 Conclusions

There are working antibodies for VMP1 and LC3 for these experiments. Mir21 mimic and mir21 inhibitor are effective at manipulating miR-21 levels in cells. And

VMP1 and NOD2 are in a protein complex which is dependent on the NACHT domain of NOD2.

4 Discussion

4.1 Sequencing

The aim of this chapter was to identify SNPs in the primary miR-21 transcript. SNPs identified can then be correlated with disease, to establish if they are relevant to IBD. They can also be correlated to methylation of the region, to establish if they can have an effect on nearby methylation. They can be correlated to expression data of nearby genes, to establish if the SNP has an effect on expression of the mRNA. And finally, functional annotation data of the region can be investigated to ascertain the likelihood of their being functional.

The sequencing that has been undertaken so far has not discovered any SNPs. The aim for this chapter was to sequence 300 patients and so far 200 have been sequenced. Additionally, the sequencing cohort is planned to be expanded by up to 600 patients from the multinational IBD-Character study. There may still be SNPs to discover in the remaining patients.

There are 335 reported SNPs in the sequenced region (Rosenbloom et al. 2015), ranging from a minor allele frequency (MAF) of 24.7% to less than 0.02%. However, the transcription start site of pri-miR-21 and the mature miR-21 sequence is highly conserved across species which may account for the low number of SNPs found in this cohort. The sequencing cohort comes from the Scottish population, in whom there is an elevated risk of IBD. Additionally, the initial evidence for altered methylation and expression of miR-21 and VMP1 in IBD was discovered in a small Scottish cohort, and has now been replicated in several other Scottish cohorts (IBD-BIOM and IBD-Character studies) as well as in a Scandinavian cohort (IBD-Character study). It is not known if the methylation findings in this region are present in populations outside of northern Europe, therefore looking for SNPs in this population is an important step in further understanding these findings.

In the optimisation stage of this experiment many more primers were tested than were finally used. Some failed to work at all, whereas others gave questionable quality sequencing data. The ones chosen for sequencing in the remaining patients gave good quality data and were designed to overlap in order to give good coverage

of the region. However, these primers had batch effects in their sequence data quality. As such, there are gaps, or areas with only one sequence read. It will be worth designing extras primers for these areas to get data that is of sufficient quality to call the sequence.

The main limitation of this experiment is the lack of good quality sequence data across the whole region. When this is completed it may still be that there are few SNPs in this cohort. Future experiments can focus on SNPs found, to establish if they are related to disease or methylation in the area. Deep sequencing is a technique that sequences many thousands of copies of the same region- this could be used to identify SNPs that are present in a small subset of cells. Cell types relevant to IBD could be investigated for SNPs in this regions.

A strength of these experiments are the data that is available for this cohort. In the 300 patients chosen there is expression data and methylation data of the region, and so if SNPs are found they can be correlated with these other data.

4.2 Pyrosequencing

Pyrosequencing is a low throughput, accessible method of determining methylation at a particular CpG site. The aim of this chapter was to develop a pyrosequencing based test which could firstly distinguish between IBD and HC, and secondly distinguish between UC and CD.

The IBD-BIOM project was a multinational EU funded project designed to look for novel biomarkers in IBD. As part of this, an epigenome wide methylation study was undertaken using Illumina 450K chips in 431 patients. These data were used to identify a selection of 8 CpG sites that could be tested in pyrosequencing in a cohort of 20 healthy controls, 20 ulcerative colitis patients and 20 Crohn's disease patients.

Two probes were chosen to distinguish between IBD and HC, which they achieved with a specificity of 82.1% and a sensitivity of 76.9% in the original data set. The remaining probes were chosen due to the differences in methylation seen in UC and CD. No probes had a statistically significant difference on their own, but in combination of several probes was hoped to be sufficient to distinguish between UC and CD.

Two of the pyrosequencing assays, CpGs in ADRA2 and ANKDD, failed to give good quality results. This is likely to be because the CpGs measured are in CpG island- regions with a high density of CpG sites- which can be difficult to quantify by pyrosequencing as sequencing primers are usually designed to not span variable positions. Pyrosequencing of the VMP1 gene has been included- although these experiments were not carried out by me, but as part of a larger study recently published (Ventham 2016). Only the data for the patients in my cohort have been shown.

RPS6KA2 and TK1 both showed significant methylation differences between IBD and HC. In both genes IBD cases were hypomethylated, and there was no difference between UC and CD. The other CpG sites did not show a statistically significant difference, but this is to be expected considering the lack of statistically significant finding in the larger cohort. There is still a difference in the median methylation and variance between IBD and HC, and this data in combination with the other probes can still provide a good model for distinguishing IBD from HC.

Linear discriminant analysis was used to create models based on these data and the performance of these models in this data set is reported. A more rigorous analysis would involve separate training and testing cohorts, and this will be performed as the next stage of these experiments, using the combination of probes that perform best in this smaller cohort. In order to perform linear discriminant analysis on a cohort with many probes complete data for every probe for each patient is required.

Pyrosequencing results were filtered based on a quality score implemented by the Qiagen pyromark software. Reasons for 'failed' quality scores include low peak heights relative to background signal, often caused by loss of sepharose beads during sample preparation using the vacuum workstation.

There were 4 CpG sites which were all present in a subset of 49 samples that were used in the final linear discriminant analysis. This produced a model that was able to distinguish IBD from HC with a sensitivity of 91.7%, and a specificity of 80%. This is an improvement on the model derived from the 450K data but as mentioned this is tested in the same data set as which the model is derived from, and its strength may be overestimated.

Being able to distinguish CD from UC would be useful clinically, and would greatly increase the utility of a methylation biomarker. In a small proportion of cases, a clear diagnosis of UC or CD can not be made (called IBD-U), and in other cases an initial diagnosis of one form of IBD may later be changed to the other form. The importance of this is that management of a patient, particularly surgical management, based on an incorrect diagnosis could have disastrous effects on that patient's outcome and quality of life.

UC and CD are similar diseases in their manifestation as well as the genetics contributing to them. Large epigenome wide methylation studies have also failed to find statistically significant differences in methylation between the two diseases, so it is unsurprising that distinguishing between them is difficult. However, it is worth investigating such models, and perhaps extending them to include a larger number of CpG sites, or to include other variables not based on methylation. It may be, with time, that this model can be used in conjuncture with other tests to diagnose IBD.

The main limitation from this experiment is that there is incomplete data for all the CpG sites for each patient, which restricts the size of the cohort. When these gaps have been filled it will be possible to test the linear discriminate analyses in all of the probes, and then select a model to validate in the larger cohort.

4.3 MiR-21 and autophagy

MiR-21 resides on the 3' end of the VMP1 gene, a key regulator of autophagy. There is some evidence already that miR-21 is involved in autophagy, and it is considered to be a key switch in the drive from a pro-inflammatory response to anti-inflammatory. There is no published data so far investigating a role for miR-21 in VMP1 expression - often microRNAs influence their host gene- although there is no target for miR-21 in the VMP1 gene. There are unpublished observations from the IBD-Character study hint at a negative correlation between miR-21 expression and VMP1 expression. Mutations in NOD2 are the strongest genetic determinant of CD, and NOD2 is known to be involved in autophagy, and mutations in the core autophagy gene ATG16L1 are also strongly associated with CD. Nod2 is involved in autophagy in a vimentin dependent manner (Stevens et al. 2014).

The aim of this chapter was to establish a working assay for measuring autophagy in cells, VMP1 levels in cells, and manipulating miR-21 levels to investigate the effect on VMP1 and autophagy. An interaction between VMP1 and NOD2 was also investigated my co-immunoprecipitation experiments.

The VMP1 antibody was optimised for western blot, and an autophagy assay, where LC3I-LC3II ratios were measured, carried out. MiR-21 inhibitor experiments and miR-21 mimic experiments were also optimised, and measured by PDCD4 mRNA levels- a target of miR-21.

A co-IP experiment was performed to establish an interaction between NOD2 and VMP1. NOD2 was successfully immunoprecipitated with VMP1 indicating they are in a larger complex. It would be worth investigating the NOD2 CD associated mutants to investigate whether this can disrupt the interaction. Knocking out other key components of the autophagy pathway may elucidate on the relationship further.

The next stages of these experiments will involve using the miR-21 inhibitor and mimic to manipulate miR-21 activity in cells and investigate the effect on VMP1. This can be used in conjunction with the autophagy assay- starvation with bafilomycin stimulation- to establish the role of miR-21 in autophagy. Finally, the addition of drugs commonly used in IBD could be used to stimulate cells and the effects on miR-21 measured.

5 Conclusions

The miR-21/VMP1 locus is a relevant locus to investigate in IBD. These experiments have established much of the groundwork in sequencing the regions, investigating methylation and expression of this region and if they correlate with SNPs. Extending previous work in paediatric IBD cases, a diagnostic model has been established and it remains to be validated in an independent cohort and the potential for distinguishing CD from UC has been investigated.

An interaction of VMP1 and NOD2 has been established by co-immunoporecipitation, further confirming NOD2's role in autophagy.

Finally, miR-21 knockdown and overexpression experiments have been optimised and therefore it will be possible to establish if it is regulator of VMP1 and what is its involvement in autophagy.

6 References

Adams, A.T. et al., 2014. Two-stage Genome-wide Methylation Profiling in Childhood-onset Crohn's Disease Implicates Epigenetic Alterations at the VMP1/MIR21 and HLA Loci. *Inflammatory bowel diseases*, 0(10), pp.1–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25144570.

Allais, L. et al., 2015. Chronic cigarette smoke exposure induces microbial and inflammatory shifts and mucin changes in the murine gut. *Environmental microbiology*, 18, pp.1352–1363. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26033517.

Ananthakrishnan, A. et al., 2013. A Prospective study of Long-term Intake of Dietary Fiber and Risk of Crohn's Disease and Ulcerative Colitis. *Gastroenterology*, 145(5), pp.970–977.

Asangani, I. a et al., 2008. MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene*, 27(15), pp.2128–36. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17968323.

Balzola, F. et al., 2010a. A human gut microbial gene catalogue established by metagenomic sequencing: Commentary. *Inflammatory Bowel Disease Monitor*, 11(1), p.28.

Balzola, F. et al., 2010b. Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry: Commentary. *Inflammatory Bowel Disease Monitor*, 10(4), pp.140–141. Available at: http://dx.doi.org/10.1038/ni.1823.

Bauckman, K.A., Owusu-Boaitey, N. & Mysorekar, I.U., 2015. Selective autophagy: Xenophagy. *Methods*, 75, pp.120–127.

Beaugerie, L. et al., 2001. Impact of cessation of smoking on the course of ulcerative colitis. *American Journal of Gastroenterology*, 96(7), pp.2113–2116.

Benchimol, E.I. et al., 2015. Inflammatory Bowel Disease in Immigrants to Canada

And Their Children: A Population-Based Cohort Study. *The American Journal of Gastroenterology*, 110(November 2014), pp.1–11. Available at: http://www.nature.com/doifinder/10.1038/ajg.2015.52.

Biedermann, L. et al., 2013. Smoking Cessation Induces Profound Changes in the Composition of the Intestinal Microbiota in Humans. *PLoS ONE*, 8(3).

Bird, A., 2002. DNA methylation patterns and epigenetic memory DNA methylation patterns and epigenetic memory. *Genes & Development*, pp.6–21.

Burisch, J. et al., 2014. Environmental factors in a population-based inception cohort of inflammatory bowel disease patients in Europe - An ECCO-EpiCom study. *Journal of Crohn's and Colitis*, 8(7), pp.607–616. Available at: http://dx.doi.org/10.1016/j.crohns.2013.11.021.

Calkins, B.M., 1989. A meta-analysis of the role of smoking in inflammatory bowel disease. *Digestive Diseases and Sciences*, 34(12), pp.1841–1854.

Cooney, R. et al., 2010. NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nature medicine*, 16(1), pp.90–97. Available at: http://dx.doi.org/10.1038/nm.2069.

Cuthbert, A.P. et al., 2002. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology*, 122(4), pp.867–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11910337 [Accessed May 18, 2012].

Deaton, A. & Bird, A., 2011. CpG islands and the regulation of transcription. *Genes & development*, 25(10), pp.1010–1022. Available at: http://genesdev.cshlp.org/content/25/10/1010.short.

Ding, X. et al., 2013. NIH Public Access. , 16(3), pp.387–393.

Eskelinen, E.-L. & Saftig, P., 2009. Autophagy: a lysosomal degradation pathway with a central role in health and disease. *Biochimica et biophysica acta*, 1793(4), pp.664–673. Available at: http://dx.doi.org/10.1016/j.bbamcr.2008.07.014.

Fenoglio, C. et al., 2011. Expression and genetic analysis of miRNAs involved in CD4+ cell activation in patients with multiple sclerosis. *Neuroscience Letters*, 504(1), pp.9–12. Available at: http://dx.doi.org/10.1016/j.neulet.2011.08.021.

Fujita, S. et al., 2008. miR-21 Gene expression triggered by AP-1 is sustained through a double-negative feedback mechanism. *Journal of molecular biology*, 378(3), pp.492–504.

Gao, Y. et al., 2013. Downregulation of BCL11A by siRNA induces apoptosis in B lymphoma cell lines. *Biomedical Reports*, 1(1), pp.47–52. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3956826/\nhttp://www.ncbi.nlm.nih.gov.gate1.inist.fr/pmc/articles/PMC3956826/pdf/br-01-01-0047.pdf.

Goujon, M. et al., 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, 38(SUPPL. 2), pp.695–699.

Grimes, C.L. et al., 2012. The innate immune protein Nod2 binds directly to MDP, a bacterial cell wall fragment. *Journal of the American Chemical Society*, 134(33), pp.13535–13537.

Halme, L. et al., 2006. Family and twin studies in inflammatory bowel disease. *World Journal of Gastroenterology*, 12(23), pp.3668–3672.

Harries, A.D., Baird, A. & Rhodes, J., 1982. Non-smoking: a feature of ulcerative colitis. *Br Med J*, 284(6317), p.706.

Higuchi, L.M. et al., 2012. A prospective study of cigarette smoking and the risk of inflammatory bowel disease in women. *The American journal of gastroenterology*, 107(9), pp.1399–406. Available at: http://dx.doi.org/10.1038/ajg.2012.196.

Hoffjan, S. et al., 2005. Gene-environment interaction effects on the development of immune responses in the 1st year of life. *American journal of human genetics*, 76(4), pp.696–704.

Hwang, J. & Weiss, R.E., 2014. Steroid-induced diabetes: a clinical and molecular approach to understanding and treatment. *Diabetes/Metabolism Research and Reviews*, 32(30), pp.96–102. Available at: http://libweb.anglia.ac.uk/.

Jakobsen, C. et al., 2013. Environmental factors and risk of developing paediatric inflammatory bowel disease - A population based study 2007-2009. *Journal of Crohn's and Colitis*, 7(1), pp.79–88. Available at: http://dx.doi.org/10.1016/j.crohns.2012.05.024.

Jongstra-Bilen, J. et al., 1992. The lymphocyte-specific protein LSP1 binds to F-actin and to the cytoskeleton through its COOH-terminal basic domain. *Journal of Cell Biology*, 118(6), pp.1443–1453.

Jostins, L.M. et al., 2012. Host microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, 490(7422), pp.119–124. Available at: http://dx.doi.org/10.1038/nature11582.

Kang, R. et al., 2011. The Beclin 1 network regulates autophagy and apoptosis. *Cell death and differentiation*, 18(4), pp.571–80. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3131912&tool=pmcentrez&rendertype=abstract.

Kaur, J. & Debnath, J., 2015. Autophagy at the crossroads of catabolism and anabolism. *Nature reviews. Molecular cell biology*, 16(8), pp.461–472. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26177004.

Khaled, W.T. et al., 2015. BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. *Nature communications*, 6, p.5987. Available at: http://www.nature.com/ncomms/2015/150109/ncomms6987/full/ncomms6987.html.

Koloski, N.A., Bret, L. & Radford-Smith, G., 2008. Hygiene hypothesis in inflammatory bowel disease: A critical review of the literature. *World Journal of Gastroenterology : WJG*, 14(2), pp.165–173.

Kumarswamy, R., Volkmann, I. & Thum, T., 2011. Regulation and function of miRNA-21 in health and disease. *RNA Biology*, 8(5), pp.706–713.

Lander, E. & Kruglyak, L., 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics*, 11(3), pp.241–247.

Available at: http://www.ncbi.nlm.nih.gov/pubmed/7581446\nhttp://www.nature.com/doifinder/10.1038/ng1195-241.

Liu, J.Z. et al., 2015. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics*, 47(9), pp.979–989. Available at: http://dx.doi.org/10.1038/ng.3359.

Liu, X. et al., 2015. MiR-21 inhibits autophagy by targeting Rab11a in renal ischemia/reperfusion. *Experimental Cell Research*, 338(1), pp.64–69. Available at: http://dx.doi.org/10.1016/j.yexcr.2015.08.010.

Lofton-Day, C. et al., 2008. DNA methylation biomarkers for blood-based colorectal cancer screening. *Clinical Chemistry*, 54(2), pp.414–423.

Matsushita, M. et al., 2005. Appendix is a priming site in the development of ulcerative colitis. *World Journal of Gastroenterology*, 11(31), pp.4869–4874.

McCarroll, S. a et al., 2008. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nature genetics*, 40(9), pp.1107–1112.

Meisgen, F. et al., 2012. miR-21 is up-regulated in psoriasis and suppresses T cell apoptosis. *Experimental Dermatology*, 21(4), pp.312–314.

Metcalf, R.A. et al., 2014. Myeloid cell nuclear differentiation antigen is expressed in a subset of marginal zone lymphomas and is useful in the differential diagnosis with follicular lymphoma. *Human Pathology*, 45(8), pp.1730–1736. Available at: http://dx.doi.org/10.1016/j.humpath.2014.04.004.

Molejon, M.I. et al., 2013. The VMP1-Beclin 1 interaction regulates autophagy induction. *Scientific reports*, 3, p.1055. Available at: http://www.nature.com/srep/2013/130111/srep01055/full/srep01055.html.

Molodecky, N.A. et al., 2012. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*, 142(1), pp.46–54.e42. Available at:

http://dx.doi.org/10.1053/j.gastro.2011.10.001.

Murthy, A. et al., 2014. A Crohn's disease variant in Atg16l1 enhances its degradation by caspase 3. *Nature*, 506(7489), pp.456–462. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24553140.

Ng, S.C., 2015. Emerging leadership lecture: Inflammatory bowel disease in Asia: Emergence of a "Western" disease. *Journal of Gastroenterology and Hepatology (Australia)*, 30(3), pp.440–445.

Nunes, T. et al., 2016. Impact of Smoking Cessation on the Clinical Course of Crohn's Disease Under Current Therapeutic Algorithms: A Multicenter Prospective Study. *The American Journal of Gastroenterology*, (November 2015), pp.1–9. Available at: http://www.nature.com/doifinder/10.1038/ajg.2015.401.

Ogura, Y. et al., 2001. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*, 411(6837), pp.603–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11385577 [Accessed May 18, 2012].

Orholm, M. et al., 1991. Familial Occurance of Inflammaotry Bowel Disease. *The New England Journal of Medicine*, 324(2), pp.84–85.

Osipova, J. et al., 2014. Diabetes-associated microRNAs in pediatric patients with type 1 diabetes mellitus: a cross-sectional cohort study. *The Journal of clinical endocrinology and metabolism*, 99(9), pp.E1661–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24937532.

Probert, C.S. et al., 1993. Prevalence and family risk of ulcerative colitis and Crohn's disease: an epidemiological study among Europeans and south Asians in Leicestershire. *Gut*, 34(11), pp.1547–51. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1374420&tool=pmcentrez&rendertype=abstract.

Raelson, J. V et al., 2007. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proceedings of the National Academy of Sciences of the United States of*

America, 104(37), pp.14747–14752. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1965486{&}tool=pmcentrez{&}rendertype=abstract.

Reese, G.E. et al., 2008. The effect of smoking after surgery for Crohn's disease: A meta-analysis of observational studies. *International Journal of Colorectal Disease*, 23(12), pp.1213–1221.

Ribas, J. et al., 2012. A novel source for miR-21 expression through the alternative polyadenylation of VMP1 gene transcripts. *Nucleic Acids Research*, 40(14), pp.6821–6833.

Roblin, X. et al., 2012. Local appendiceal dysbiosis: the missing link between the appendix and ulcerative colitis? *Gut*, 61(4), pp.635–636.

Ropolo, A. et al., 2007. The pancreatitis-induced vacuole membrane protein 1 triggers autophagy in mammalian cells. *Journal of Biological Chemistry*, 282(51), pp.37124–37133.

Rosenbloom, K.R. et al., 2015. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research*, 43(D1), pp.D670–D681.

Russel, M.G. et al., 1997. Appendectomy and the Risk of Developing Ulcerative Colitis or Crohn's Disease: Results of a Large Case-Control Study. *Gastroenterology*, 113(2), pp.377–382.

Russell, R.C., Yuan, H.-X. & Guan, K.-L., 2014. Autophagy regulation by nutrient signaling. *Cell research*, 24(1), pp.42–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24343578.

Ryan, N.M. et al., 2014. SuRFing the genomics wave: an R package for prioritising SNPs by functionality. *Genome Medicine*, 6, pp.1–13.

Sha, M. et al., 2015. Celastrol induces cell cycle arrest by MicroRNA-21-mTOR-mediated inhibition p27 protein degradation in gastric cancer. *Cancer Cell International*, 15(1), p.101. Available at: http://www.cancerci.com/content/15/1/101.

Sheedy, F.J., 2015. Turning 21: Induction of miR-21 as a key switch in the inflammatory response. *Frontiers in Immunology*, 6(JAN), pp.1–9.

Stevens, C. et al., 2014. Europe PMC Funders Group The intermediate filament protein vimentin is a regulator of NOD2 activity. , 62(5), pp.695–707.

Szamosi, T. et al., 2010. Early azathioprine/biological therapy is associated with decreased risk for first surgery and delays time to surgery but not reoperation in both smokers and nonsmokers with Crohn's disease, while smoking decreases the risk of colectomy in ulcerative colit. *European journal of gastroenterology & hepatology*, 22(7), pp.872–9. Available at: http://ep.fjernadgang.kb.dk/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed9&AN=2010357388\nhttp://www.ncbi.nlm.nih.gov/pubmed/19648821.

Thaiss, C.A. et al., 2016. The microbiome and innate immunity. *Nature*, 535(7610), pp.65–74. Available at: http://www.nature.com/doifinder/10.1038/nature18847.

Toiyama, Y. et al., 2013. Serum miR-21 as a diagnostic and prognostic biomarker in colorectal cancer. *Journal of the National Cancer Institute*, 105(12), pp.849–859.

Tooze, S. a & Yoshimori, T., 2010. The origin of the autophagosomal membrane. *Nature cell biology*, 12(9), pp.831–835. Available at: http://dx.doi.org/10.1038/ncb0910-831.

Venables, W. N. & Ripley, B.D., 2002. *Modern Applied Statistics with S.*,

Ventham, N.T., 2016. Integrative genome-wide analysis demonstrates DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nature communications*, p.In Press.

Winter, J. et al., 2009. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature cell biology*, 11(3), pp.228–234.

Wu, H. et al., 2007. miRNA profiling of na??ve, effector and memory CD8 T cells. *PLoS ONE*, 2(10).

Zeilinger, S. et al., 2013. Tobacco Smoking Leads to Extensive Genome-Wide Changes in DNA Methylation. *PLoS ONE*, 8(5).